Journal of Machine Learning Research 23 (2022) 1-50 Submitted 9/20; Revised 4/22; Published 10/22

Community detection in sparse latent space models

Fengnan Gao FNGAO@FUDAN.EDU.CN
School of Data Science, Fudan University and
Shanghai Center for Mathematical Sciences
N202 Zibin, 220 Handan Road, Shanghai 200433, China

Zongming Ma ZONGMING@WHARTON.UPENN.EDU
Department of Statistics and Data Science
University of Pennsylvania
265 South 37th Street, Philadelphia, PA 19104, USA

Hongsong Yuan YUAN.HONGSONG@SHUFE.EDU.CN
Research Institute for Interdisciplinary Sciences and
School of Information Management and Engineering
Shanghai University of Finance and Economics
777 Guoding Road, Shanghai 200433, China

Editor: Edo Airoldi

Abstract

We show that a simple community detection algorithm originated from stochastic blockmodel literature achieves consistency, and even optimality, for a broad and flexible class of sparse latent space models. The class of models includes latent eigenmodels (Hoff, 2008). The community detection algorithm is based on spectral clustering followed by local refinement via normalized edge counting. It is easy to implement and attains high accuracy with a low computational budget. The proof of its optimality depends on a neat equivalence between likelihood ratio test and edge counting in a simple vs. simple hypothesis testing problem that underpins the refinement step, which could be of independent interest.

Keywords: blockmodel, eigenmodel, minimax rates, social network, spectral clustering

1. Introduction

Network is a prevalent form of relational data. A central theme in learning network data is community detection (Goldenberg et al., 2010; Fortunato, 2010). Community detection seeks to partition the nodes of a network into several disjoint subsets (a.k.a. communities) upon observing the adjacency matrix (Girvan and Newman, 2002). The underlying assumption is that nodes within the same community share some commonalities in their connection patterns. To understand and to motivate algorithms for community detection, statisticians, probabilists and theoretical computer scientists have studied stochastic blockmodels (SBMs, Holland et al., 1983) extensively. To date, researchers have obtained a thorough understanding of the fundamental limits and the behavior of various algorithms under SBMs. For more details, we refer interested readers to the review papers Abbe (2017); Moore (2017) and the references therein. A major shortcoming of SBMs is that nodes within the same community must have exactly the same degree profile, and hence SBMs cannot model degree heterogeneity which is commonly observed in real world networks. To mitigate this issue, researchers have proposed degree-corrected blockmodels (DCBMs) where an extra sequence of degree correction parameters was used to lend more flexibility to individual node degrees (Karrer and Newman, 2011). In the regimes of strong consistency (when perfect recovery of community structure is possible) and weak consistency (when perfect recovery except for a vanishing proportion of nodes is possible), it is known that spectral clustering followed by certain local algorithm could achieve the best possible accuracy (Abbe, 2017; Gao and Ma, 2020).

In a separate line of literature, statisticians have proposed and studied a class of network models called latent space models (Hoff et al., 2002; Hoff, 2003; Handcock et al., 2007; Hoff, 2008; Krivitsky et al., 2009; Shalizi and Asta, 2021). We may view this class of models as a natural extension of generalized linear models to network setting. In this paper, we consider the following generative model for entries of the observed adjacency matrix A. For any positive integer m, let $[m] = \{1, \ldots, m\}$. First, we exclude self-loops and so $A_{ii} = 0$ for all $i \in [n]$. In addition, conditional on unobserved values of $\{\alpha_i\}_{i=1}^n$ and $\{z_i\}_{i=1}^n$, we assume that the Bernoulli random variables $\{A_{ij} = A_{ji} : 1 \leqslant i < j \leqslant n\}$ are mutually independent, and for each pair $i < j$,

$$\begin{aligned} P_{ij} &= \mathbb{P}(A_{ij} = 1 \mid \{\alpha_i, z_i\}_{i=1}^n) = 1 - \mathbb{P}(A_{ij} = 0 \mid \{\alpha_i, z_i\}_{i=1}^n) \\ &= \frac{\exp(\alpha_i + \alpha_j + z_i^\top H z_j)}{1 + \exp(\alpha_i + \alpha_j + z_i^\top H z_j)}. \end{aligned} \tag{1}$$

Model (1) is a generalization of the logistic regression model to the binary network setting. Here $\{\alpha_i\}_{i=1}^n$ is a sequence of degree parameters. Nodes with larger values of α_i's are expected to have larger degrees. Furthermore, $\{z_i\}_{i=1}^n \subset \mathbb{R}^d$ are the latent positions of the nodes in a d-dimensional latent space (a.k.a. "social space" in the latent space model literature), and H an unobserved $d \times d$ symmetric matrix that moderates how the latent positions affect edge formation. To impose a community structure, let there be k communities. Let $\{\mathcal{L}_{z,1}, \ldots, \mathcal{L}_{z,k}\}$ be k different probability distributions defined on the latent space $\mathbb{R}^d$. We assume that there is an unknown deterministic community label vector $\sigma = (\sigma_1, \ldots, \sigma_n)^\top \in [k]^n$. For each node i, $\sigma_i = j$ means the ith node belongs to the jth community. In this case z_i is a random vector generated from $\mathcal{L}_{z,\sigma_i}$, and all the z_i's are mutually independent. Our goal is to infer σ from the observed adjacency matrix A.

The latent space model (1) not only models community structures but is also flexible for modeling degree heterogeneity. The particular form (1) can be identified as the latent eigenmodel in Hoff (2008) which was shown to possess more flexibility and modeling power than many other latent space models and various blockmodels. In particular, degree-corrected blockmodel reduces to a special case of model (1) with $\{\mathcal{L}_{z,j} : 1 \leqslant j \leqslant k\}$ each assigning probability one on a distinct point. Ma et al. (2020) studied fitting methods for this model when H is the identity matrix and α_i's and z_i's are considered deterministic. See also Wu et al. (2017). Their study also revealed appealing numerical properties for clustering estimated latent positions after fitting such a special case of (1), which has partially motivated the study reported in this manuscript. Nevertheless, to the best of our limited knowledge, the literature of community detection for latent space models has been scarce. A sound understanding of community detection is crucial to applications of such models, as it provides theoretical foundations to community discoveries in modeling real-world networks with latent space models. The present manuscript aims to take a first step along this direction.

1.1 Main contributions

The main contributions of this manuscript are twofold.

- From an algorithmic viewpoint, we establish consistency of SpecLoRe, a simple and intuitive community detection method for latent space model (1) in a stylized setting. The method is based on spectral clustering followed by a local edge counting refinement step. It was first proposed for blockmodels and its properties for the broader class of latent space models, especially in the generality of latent eigenmodels, were previously unknown. Our new consistency result suggests that the method enjoys a certain level of universal applicability on exchangeable network models.

  The community detection method aims only at estimating community structure while not trying to find estimates of latent positions or their distributions. Thus, it is different in nature from most algorithms developed for latent space models in the literature which fit specific latent space models and estimate model parameters. See, for instance, Ma et al. (2020),Wu et al. (2017), and Zhang et al. (2018). As estimation of latent positions usually involves solving a computationally expensive optimization problem, our method bypasses it and attains comparable or even better accuracy for community detection with considerably lower computational cost.

- From a theoretical viewpoint, our consistency result sheds light on a better understanding of community detection for latent space models. Our explicit upper bounds on rates of convergence exhibit an interesting interplay between signal-to-noise ratio affected by network sparsity and that affected by latent positions and the quadratic form matrix H in (1). In a more restrictive setting, we could even show that the resulting estimator achieves nearly optimal rates of convergence in some minimax sense.

  The key insight comes from the investigation of a special simple vs. simple hypothesis testing problem which underpins the local refinement step in our method. We study error rates of a simple edge counting procedure for this testing problem. By a seemingly intuitive yet elegant exploitation of symmetry inherent to our model, we are able to show that the simple testing method is equivalent to the optimal likelihood ratio test under mild assumptions. The equivalence, being the major novelty of our manuscript, paves the way for establishing optimality of our algorithm.

1.2 Relation to prior work

The present manuscript is connected to Ma et al. (2020) and Wu et al. (2017) which studied efficient fitting methods for model (1) when the z_i's are treated as deterministic. Ma et al. (2020) also touched community detection for (1). However, the method was a "plug-in" one which ran k-means clustering on estimated latent positions. As we shall show empirically, its computational efficiency is far inferior to the method we consider in this paper while community detection accuracies are comparable.

Moreover, Handcock et al. (2007) and Krivitsky and Handcock (2008) proposed Bayesian algorithms for community detection in a latent distance model which is different from (1) but can be approximated by it (Ma et al., 2020). Their study emphasized the algorithmic and computational perspective, and theoretical properties of the proposed methods were not considered.

In addition to the community detection literature for blockmodels that we have mentioned earlier, there have been extensive studies of community detection for random dot-product graph models, especially via spectral methods. See the review papers Athreya et al. (2017) and the references therein. These models relax SBMs and their variants such as DCBMs and mixed membership blockmodels. However, these studies have also mostly focused on "plug-in" methods and community detection is conducted through clustering estimated latent positions. There has been little investigation on methods designed specifically for community detection, and there is little understanding on fundamental limits of such an inference goal.

1.3 Organization of paper

The rest of the manuscript is organized as follows. Section 2 presents the method and a variant of it for which we shall establish theoretical results. Section 3 states all theoretical results in an idealized setting for model (1) and the method. We demonstrate numerical prowess of the method on simulated and real data examples in Sections 4 and 5, respectively. After a brief discussion in Section 6, the appendices present detailed proofs of theoretical results.

1.4 Notation

Let $S(\cdot)$ be the sigmoid function $S : x \mapsto 1/(1+e^{-x})$, which is the inverse of the logit function $p \mapsto \log\{p/(1-p)\}$. Let $1(E)$ be the indicator function of E, where E may be an event or a set. Recall $[m] := \{1, \ldots, m\}$ and S_2 contains the two permutations of [2]. $\|A\|_2$ is the usual operator norm of A: $\|A\|_2 = \sup_{x\neq 0} \|Ax\|_2/\|x\|_2$. The Frobenius norm $\|A\|_{\mathrm{F}}$ of matrix $A = (A_{ij})_{i\in[n],j\in[m]}$ is defined as $\|A\|_{\mathrm{F}} = (\sum_i \sum_j A_{ij}^2)^{1/2}$. For vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, $\|v\|_p = \left(\sum_{i=1}^d |v_i|^p\right)^{1/p}$ for $p = 1, 2$. 1_d and 0_d denote a d-dimensional column vector with all entries equal to 1 and 0, respectively. For notational simplicity in asymptotics, for two deterministic sequences a_n and b_n, we define the following notations: $a_n \lesssim (\gtrsim)\, b_n$ if and only if there exists a constant $C > 0$ such that $a_n \leqslant (\geqslant)\, Cb_n$; $a_n \ll (\gg)\, b_n$ if and only if $a_n/b_n \to 0\,(\infty)$ as $n \to \infty$. We also write $a_n = O(b_n)$ when $a_n \lesssim b_n$, and $a_n = o(b_n)$ when $a_n \ll b_n$.

2. Method

We consider a two-stage procedure, consisting of an initialization stage and a refinement stage. The algorithm was first proposed in Gao et al. (2018) as a community detection method for DCBMs. In what follows, we introduce the two stages separately for self-completeness.

2.1 A practical version

We first introduce a practical version of our method which we shall refer to as SpecLoRe (spectral clustering followed by local refinement) in the rest of this paper. It is obtained by running Algorithm 2 with initial value given by Algorithm 1. It relies on Algorithm 1 to process the adjacency matrix for an initial guess $\widehat{\sigma}^0$ and on Algorithm 2 to further refine the crude yet informative initial guess to obtain the final estimator. Here and after, we assume the number of communities k is known.

Initialization We summarize the initialization stage as Algorithm 1. In this stage, we first compute the best rank-k approximation $\widehat{P}$ to the observed adjacency matrix A where k is the number of clusters. Note that this is easily achieved by the celebrated singular value decomposition. Then we apply weighted k-median clustering on normalized rows of $\widehat{P}$. While running weighted k-median clustering, we only seek a constant-factor approximation solution to ensure that the output could be produced within polynomial time complexity (Charikar et al., 2002; Chen et al., 2018). Here ε is required to be an absolute constant.

Algorithm 1: Initialization

1: **Input:** Adjacency matrix: A; latent dimension d; number of clusters k.

2: Find the solution to the following optimization problem

$$\widehat{P} = \mathop{\arg\min}_{\mathsf{rank}(P)\leqslant k} \|A - P\|_{\mathrm{F}}^2. \tag{2}$$

3: Let $\widehat{P}_i$ be the ith row. Define $J_0 = \{i \in [n] \mid \|\widehat{P}_i\|_1 = 0\}$. For $i \in J_0^c$, define $\widetilde{P}_i = \widehat{P}_i/\|\widehat{P}_i\|_1$. Put $\widehat{\sigma}_i^0 = 0$ for $i \in J_0$.

4: Find a $(1+\varepsilon)$ approximate weighted k-median solution for clustering $(\widetilde{P}_i)_{i=1}^n$. That is, find labels $\widehat{\sigma}^0 = \{\widehat{\sigma}_i^0\}_{i=1}^n \in [k]^n$ and centers $\widehat{v}_l \in \mathbb{R}^k, l = 1, \cdots, k$, such that

$$\sum_{l=1}^k \min_{v_l \in \mathbb{R}^n} \sum_{\{i \in J_0^c : \hat{\sigma}_i^0 = l\}} \|\widehat{P}_i\|_1 \|\widetilde{P}_i - \widehat{v}_l\|_1 \leqslant (1+\varepsilon) \min_{\sigma \in [k]^n} \sum_{l=1}^k \min_{v_l \in \mathbb{R}^k} \sum_{\{i : \sigma_i = l\}} \|\widehat{P}_i\|_1 \|\widetilde{P}_i - v_l\|_1.$$

5: **Output:** $\widehat{\sigma}^0$.

Refinement We then state the local refinement procedure in Algorithm 2. Starting with an initial estimator $\widehat{\sigma}^0$, we refine it by the following simple and intuitive majority voting rule. For node i, we look at all communities prescribed in $\widehat{\sigma}^0$ and calculate the relative connecting frequency from i to each community. Then we recalibrate the community label of node i to be that of the community to which it most likely connects. Since the refinement is strictly local, it can be easily carried out in a parallel fashion on each node. As the process only involves counting edges, a crude inspection of the algorithm puts the computational cost of one round of refinement at $O(n^2)$. Moreover, as simulated and real world examples reported in Sections 4 and 5 suggest, one typically only needs to run an $O(1)$ round of refinement to arrive at a stable estimator.

Algorithm 2: Local Refinement

Input: Adjacency matrix: A; number of clusters k; an initial label vector $\widehat{\sigma}^0$; number of iterations R.
Initialize $\widehat{\sigma}^{\text{old}} := \widehat{\sigma}^0$.
for $t \leftarrow 1$ **to** R **do**
  for $i \leftarrow 1$ **to** n **do**
    Update the labels

$$\widehat{\sigma}_i^{\text{new}} := \arg\max_{u \in [k]} \frac{1}{|\{j : \widehat{\sigma}_j^{\text{old}} = u\}|} \sum_{\{j : \widehat{\sigma}_j^{\text{old}} = u\}} A_{ij}.$$

  end for
  $\widehat{\sigma}^{\text{old}} := \widehat{\sigma}^{\text{new}}$.
end for
Output: $\widehat{\sigma} := \widehat{\sigma}^{\text{new}}$.

2.2 A theoretically justifiable variant

In this part, we state a theoretically justifiable variant of SpecLoRe, summarized as Algorithm 3, for which we will establish an upper bound in Section 3. As an artifact of our proof techniques (see the proof of Theorem 8), we are unable to present a cleaner theory for SpecLoRe. As a remedy, the new comprehensive Algorithm 3 has two stages as well and combines both Algorithms 1 and 2, albeit not in a simple consecutive fashion.

The first part of Algorithm 3 (lines 2–7) does a separate initialization on each node by performing Algorithm 1 on the network excluding node i, leading to a vector $\widehat{\sigma}^{(-i,0)}$. It then applies Algorithm 2 on $\widehat{\sigma}^{(-i,0)}$ to obtain a refined estimate for node i, denoted by $\widehat{\sigma}_i^{(-i,0)}$. The separate initializations dissolve an issue in the proof. However, since each initialization could end up with a different permutation of community labels, the second part of Algorithm 3 (lines 8–11) aligns all label permutations with that of $\widehat{\sigma}^{(-1,0)}$.

Algorithm 3 has at most polynomial time complexity. We do not emphasize its computational efficiency though, since we view it more as a proof device rather than a practical replacement of SpecLoRe in the previous subsection.

3. Theoretical results

We present decision theoretic results for Algorithm 3 on model (1). We focus on the balanced two community case, i.e., we consider the case where $k = 2$ and the two communities have roughly equal sizes. The need to consider Algorithm 3 is due to proof technique, and we show in later sections that there is little numerical difference between its accuracy and that of SpecLoRe in Section 2.1.

3.1 A decision-theoretic framework

We shall establish uniform high probability error bounds for Algorithm 3. To this end, we first define classes of models for which uniform error bounds are to be obtained.

Algorithm 3: A provable version of latent space model community detection method

Input: Adjacency matrix: A; latent dimension d; number of clusters k.
for $i \leftarrow 1$ **to** n **do**
  Let $A^{(-i)} \in \{0,1\}^{(n-1)\times(n-1)}$ be the matrix obtained from removing the ith row and the ith column of A;
  Apply Algorithm 1 on $A^{(-i)}$ to obtain $\widehat{\sigma}^{(-i,0)} \in [k]^{n-1}$;
  Augment $\widehat{\sigma}^{(-i,0)}$ to an n-dimensional vector by inserting 0 in the ith position;
  Update

$$\widehat{\sigma}_i^{(-i,0)} = \arg\max_{u\in[k]} \frac{1}{|\{j : \widehat{\sigma}_j^{(-i,0)} = u\}|} \sum_{\{j:\widehat{\sigma}_j^{(-i,0)}=u\}} A_{ij}.$$

end for
Define $\widehat{\sigma}_1 = \widehat{\sigma}_1^{(-1,0)}$.
for $i \leftarrow 2$ **to** n **do**
  Let

$$\widehat{\sigma}_i = \arg\max_{u\in[k]} \left|\{j : \widehat{\sigma}_j^{(-1,0)} = u\} \cap \{j : \widehat{\sigma}_j^{(-i,0)} = \widehat{\sigma}_i^{(-i,0)}\}\right|.$$

end for
Output: $\widehat{\sigma} = (\widehat{\sigma}_1, \dots, \widehat{\sigma}_n)^\top \in [k]^n$.

Uniformity class Let the adjacency matrix be $A = (A_{ij}) = A^\top \in \{0,1\}^{n\times n}$. Given a *deterministic* community label vector $\sigma \in [2]^n$, we suppose that the edges are generated in the following way:

$$\begin{aligned}
&\alpha_i \overset{iid}{\sim} F_\alpha, \quad z_i \overset{ind}{\sim} F_{z,\sigma_i}, \quad i \in [n], \\
&A_{ij} = A_{ji} \mid \alpha_i, \alpha_j, z_i, z_j \overset{ind}{\sim} \text{Bernoulli}(P_{ij}), \quad i, j \in [n], \\
&\text{logit}(P_{ij}) = \alpha_i + \alpha_j + z_i^\top H z_j.
\end{aligned} \tag{3}$$

Here F_α is a distribution from which the α_i's are generated, and H is a symmetric $n \times n$ matrix. The two distributions $\{F_{z,j} : j = 1, 2\}$ generate each latent position z_i depending on the value of σ_i. For most of theoretical results below, we further assume that

$$F_{z,j} \overset{d}{=} N_d((-1)^{j-1}\mu, \tau^2 I_d), \quad j = 1, 2. \tag{4}$$

In other words, we assume that the latent positions within each community are generated according to an isotropic multivariate Gaussian distribution with shared covariance structure and different mean vector depending on the community label[1]. Here and after, I_d is the $d \times d$ identity matrix. For identifiability of μ, τ and H, we assume that

$$\|H\|_2 = 1. \tag{5}$$

1. In view of Lemma 2 later, the same lower and upper bounds hold if the two component mean vectors are in general positions μ_1 and μ_2 with $\|\mu_1\|_2 = \|\mu_2\|_2$ instead of being symmetric about origin. We choose the symmetric version mainly for convenience of arguments.

In what follows, we denote such a model by $\mathcal{M}_n(\sigma, H, \mu, \tau, F_\alpha)$. For each $\sigma \in [2]^n$ and each $j \in [2]$, let $n_j = n_j(\sigma) = |\{i : \sigma_i = j\}|$. The uniformity classes of interest are of the form

$$\mathcal{P}_n(H, \mu, \tau, F_\alpha) = \left\{ \mathcal{M}_n(\sigma, H, \mu, \tau, F_\alpha) : n_j(\sigma) \in \left[(1-\delta_n)\frac{n}{2}, (1+\delta_n)\frac{n}{2}\right], \quad j = 1, 2 \right\}, \quad (6)$$

where $\delta_n = o(1)$ is some vanishing sequence. In the rest of this section, we treat H and μ as fixed parameters, while τ and F_α scale with n.

Estimation and loss function Our goal is to estimate the community labels $\{\sigma_i : i \in [n]\}$ based on the observed adjacency matrix A. Since permutation of community labels does not change the partition of nodes, we use the following misclustering proportion as the loss function

$$\ell(\sigma, \widehat{\sigma}) = \min_{\pi \in S_2} \frac{1}{n} \sum_{i=1}^{n} 1(\widehat{\sigma}_i \neq \pi(\sigma_i)). \quad (7)$$

3.2 Assumptions on model parameters

For convenience of reference, we collect and explain various assumptions used in main results here.

Assumption 1 *For $i \in [n]$, $\alpha_i = \overline{\alpha} + \omega_i$, with $\overline{\alpha}$ deterministic, ω_i i.i.d. with $\mathbb{E}(\omega_1) = 0$, $\mathbb{E}(e^{2\omega_1}) \leqslant C$ for some constant $C > 0$, and*

$$-\underline{\omega} \leqslant \omega_i \leqslant \omega', \quad (8)$$

where $\underline{\omega} > 0$ is a constant but ω' is allowed to grow to ∞ with n. As $n \to \infty$, $\overline{\alpha}$ and ω' jointly satisfy all the following conditions

$$\overline{\alpha} + \omega' \to -\infty, \quad (9)$$

$$ne^{2\overline{\alpha}}/(\log n)^{1/2} \to \infty, \quad (10)$$

$$e^{\omega'} / \min\left\{ne^{2\overline{\alpha}}, n/\log n\right\} \to 0. \quad (11)$$

Furthermore, for some constants $\overline{L} > 0$ and $C_1 > 0$, the empirical fourth moment of e^{ω_i} satisfies the condition

$$\mathbb{P}\left\{ \Big(\frac{1}{n_u} \sum_{\sigma_i = u} e^{4\omega_i}\Big)^{1/4} > \overline{L} \right\} \leqslant n^{-(1+C_1)}, \quad \text{for } u \in [2]. \quad (12)$$

In this overarching assumption on F_α, equation (9) ensures that the network is sparse in the sense that the maximum degree scales at an $o(n)$ rate. Equations (8) and (10) jointly imply that the minimum degree grows at a rate no slower than $(\log n)^{1/2}$. Equation (11) guarantees that the maximum degree grows at a slower rate than squared minimum degree. Moreover, it imposes the restriction that the ratio of maximum over minimum degrees grows at a slower rate than $n/\log n$. Finally, (12) puts some technical tail bounds on the empirical fourth moments of e^{ω_i} within each community.

Assumption 2 *There exists a positive constant c such that $\tau(\log n)^{1/2} \leqslant c$.*

Even if we directly observe the latent positions $\{z_i\}_{i=1}^n$, we always suffer the Bayes error for clustering two normal distributions with identical covariance structure. Write $\bar{\Phi}(t) = \mathbb{P}\{N(0,1) \geqslant t\}$. Under model (3)–(4), simple calculation shows that the Bayes error is at the rate $\bar{\Phi}(\|\mu\|_2/\tau) \lesssim \exp\{-\|\mu\|_2^2/(2\tau^2)\}\tau/\|\mu\|_2$ as $n \to \infty$. Since μ is fixed, by varying c, Assumption 2 allows us to consider any case where the Bayes error scales at an $O(n^{-a})$ rate for any $a > 0$.

Assumption 3 *For H in (1) and μ in (4), $\mu^\top H\mu > 0$.*

This is an assortativity assumption. With this assumption, we make certain that, given the same α_i values, nodes within the same community are more likely to be connected than nodes from two different communities. It can hold even when H is not positive semi-definite.

Assumption 4 *For H in (1) and μ in (4), μ is an eigenvector of H associated with some positive eigenvalue.*

This assumption is a strengthened version of Assumption 3. It is trivially true when $H = I_d$ is the identity matrix. We only need this assumption when minimax lower bounds are concerned.

Remark 1 *We take the following simple example to see what Assumption 4 entails. Let $H = diag(1_{d_1}^\top, -1_{d-d_1}^\top)$. The inner product defined by H results in $P_{ij} = S(\alpha_i + \alpha_j + z_i^{(1)} z_j^{(1)} - z_i^{(2)} z_j^{(2)})$, where the superscript (1) and (2) indicate the vector made of the first d_1 coordinates and the last $d - d_1$ coordinates of z, respectively. Possible μ's, allowing the above argument to work, can take value in the d_1-dim. subspace such as $\mu = \left(\left(\mu^{(1)}\right)^\top, 0_{d-d_1}^\top\right)^\top$. This means the latent variable z can be decomposed into two components, the signal component $z^{(1)}$ and the noise component $z^{(2)}$,*

$$z = \begin{pmatrix} z^{(1)} \sim \mu^{(1)} + N_{d_1}(0, I_{d_1}) \\ z^{(2)} \sim N_{d-d_1}(0, I_{d-d_1}) \end{pmatrix}.$$

The signal component enhances the clustering and the noise reduces signal-to-noise ratio. In effect, this allows some additional flexibility in adding some noise in the latent variable.

3.3 A closely related testing problem

We first consider the following testing problem, which applies to slightly more general settings than the model setup that we usually take in the rest of the manuscript.

Suppose that we observe a network of size $2m + 1$, with m nodes $1, \ldots, m$ having known labels 1 and m nodes $m + 1, \ldots, 2m$ having labels 2. Suppose that node 0 has the only unknown label σ_0. Further, assume that we have some base distribution F with density f and write F_ν as its shifted version by ν with density f_ν, i.e., $f_\nu(z) = f(z - \nu)$. In addition, we assume that for nodes in the first community, z_i are i.i.d. and follow distribution F_{μ_1} and for those in the second, z_i are i.i.d. and follow distribution F_{μ_2}. We proceed to consider testing the following hypotheses

$$H_0 : \sigma_0 = 1, \quad \text{versus} \quad H_1 : \sigma_0 = 2. \tag{13}$$

Let $A_{0,i} = 1$ if there is an edge between nodes 0 and i, and otherwise 0. Under our modeling assumption, conditional on the realization of the α's and the z's, $\{A_{0,i} : i = 1, \ldots, 2m\}$ are independent Bernoulli random variables with success probability $P_{0i} = S(z_0^\top H z_i + \alpha_0 + \alpha_i)$. Define $A_0^{(1)} = \sum_{i=1}^m A_{0,i}$ and $A_0^{(2)} = \sum_{i=m+1}^{2m} A_{0,i}$.

3.3.1 Likelihood ratio test and edge counting

The following lemma connects the likelihood ratio test for (13) and edge counting. The proof hinges on symmetry in likelihoods under null and alternative hypotheses, which in turn results from the fact that integrating over symmetric latent distributions keeps the symmetry in place. We keep its proof in the main text as it may cast light on models of a larger class, when capitalizing on similar symmetries in latent distributions is possible.

For any $z \in \mathbb{R}^d$, define the Householder reflection mapping about the hyperplane $\{z : z^\top v = 0\}$ for some unit vector v by $z \mapsto z^m$ where $z^m = (I_d - 2vv^\top)z$. Note $(z^m)^m = z$.

Lemma 2 *Consider the hypothesis testing problem* (13). *Suppose* $\|\mu_1\|_2 = \|\mu_2\|_2$. *Suppose that* f *satisfies that* $f(z_1) = f(z_2)$ *for* $\|z_1\|_2 = \|z_2\|_2$ *and that* $f_{\mu_1}(z) > f_{\mu_2}(z)$ *on* $\{z : z^\top(\mu_1 - \mu_2) > 0\}$. *Suppose that* H *satisfies* $z_1^\top H z_2 = (z_1^m)^\top H z_2^m$ *for all* z_1 *and* z_2 *and* $\{z : z^\top H(\mu_1 - \mu_2) > 0\} = \{z : z^\top(\mu_1 - \mu_2) > 0\}$. *Suppose that* $\{\alpha_i : 0 \leqslant i \leqslant n\}$ *are i.i.d. Then the likelihood ratio test which reject* H_1 *when the likelihood ratio of alternative over null is larger than* 1 *is equivalent to the simple edge counting test where we reject* H_1 *when* $A_0^{(1)} < A_0^{(2)}$.

Proof Define $v := (\mu_1 - \mu_2)/\|\mu_1 - \mu_2\|_2$. Since $\|\mu_1\|_2 = \|\mu_2\|_2$, $\{z : \big(z - (\mu_1 + \mu_2)/2\big)^\top(\mu_1 - \mu_2) = 0\} = \{z : z^\top v = 0\}$, whence we may define the Householder transformation $z \mapsto z^m$ by $z^m = (I - 2vv^\top)z$. Note that $\|z - \mu_1\|_2 = \|z^m - \mu_2\|_2$.

To simplify notation, write $F_1(\cdot)$ and $F_2(\cdot)$ as shorthands of F_{μ_1} and F_{μ_2}, respectively, and $f_1(\cdot)$ and $f_2(\cdot)$ the corresponding densities. Let F_α be the distribution of α's. Define the following quantities

$$p(\alpha_0, z_0) = \iint S(z_0^\top H z + \alpha_0 + \alpha) dF_\alpha(\alpha) dF_1(z), \tag{14}$$

$$q(\alpha_0, z_0) = \iint S(z_0^\top H z + \alpha_0 + \alpha) dF_\alpha(\alpha) dF_2(z). \tag{15}$$

Noticing that $dF_1(z) = f(z - \mu_1)dz = f(z^m - \mu_2)dz = dF_2(z^m)$ as $\|z - \mu_1\|_2 = \|z^m - \mu_2\|_2$, we have by assumption that

$$\begin{aligned} q(\alpha_0, z_0) &= \iint S((z_0^m)^\top H z^m + \alpha_0 + \alpha) dF_\alpha(\alpha) dF_2(z) \\ &= \iint S((z_0^m)^\top H z^m + \alpha_0 + \alpha) dF_\alpha(\alpha) dF_1(z^m) = p(\alpha_0, z_0^m). \end{aligned}$$

The first equality holds since $(z_0^m)^\top H z^m = z_0^\top H z$ for all z_0 and z. Conditioned on z_0 and α_0, by Fubini's theorem, we obtain the conditional likelihood

$$g(\alpha_0, z_0) = \big\{p(\alpha_0, z_0)\big\}^{A_0^{(1)}} \big\{1 - p(\alpha_0, z_0)\big\}^{m - A_0^{(1)}} \big\{q(\alpha_0, z_0)\big\}^{A_0^{(2)}} \big\{1 - q(\alpha_0, z_0)\big\}^{m - A_0^{(2)}}.$$

We obtain $g(\alpha_0, z_0^m)$ by plugging in z_0^m in the preceding display and noticing $p(\alpha_0, z_0^m) = q(\alpha_0, z_0)$

$$g(\alpha_0, z_0^m) = \big\{q(\alpha_0, z_0)\big\}^{A_0^{(1)}} \big\{1 - q(\alpha_0, z_0)\big\}^{m - A_0^{(1)}} \big\{p(\alpha_0, z_0)\big\}^{A_0^{(2)}} \big\{1 - p(\alpha_0, z_0)\big\}^{m - A_0^{(2)}}.$$

The full likelihood under H_0, denoted by I_1 (as $\sigma_0 = 1$), minus the full likelihood under H_1, I_2 (as $\sigma_0 = 2$), is

$$\begin{aligned} I_1 - I_2 &= \iint g(\alpha_0, z_0) dF_\alpha(\alpha_0) dF_1(z_0) - \iint g(\alpha_0, z_0) dF_\alpha(\alpha_0) dF_2(z_0) \\ &= \int \left[\int \{g(\alpha_0, z_0) - g(\alpha_0, z_0^m)\} dF_\alpha(\alpha_0) \right] dF_1(z_0). \end{aligned}$$

We define the above integrand inside the square brackets to be $G(z_0)$ and write p and q as shorthands of $q(\alpha_0, z_0)$ and $p(\alpha_0, z_0)$, respectively. So

$$\begin{aligned} &G(z_0) \\ &= \int \left[(1-p)^m (1-q)^m \left\{ \left(\frac{p}{1-p}\right)^{A_0^{(1)}} \left(\frac{q}{1-q}\right)^{A_0^{(2)}} - \left(\frac{q}{1-q}\right)^{A_0^{(1)}} \left(\frac{p}{1-p}\right)^{A_0^{(2)}} \right\} \right] dF_\alpha(\alpha_0). \end{aligned}$$

Moreover, since $p(\alpha_0, z_0^m) = q(\alpha_0, z_0)$, we have $G(z_0^m) = -G(z_0)$. If $A_0^{(1)} = A_0^{(2)}$, the preceding display is 0 and $I_1 = I_2$, whence we may not differentiate between H_0 and H_1. For the rest of this proof, we consider $A_0^{(1)} > A_0^{(2)}$.

Define $\mathcal{L}_1 := \{z : z^\top H(\mu_1 - \mu_2) > 0\}$ and $\mathcal{L}_2 := \{z : z^\top H(\mu_1 - \mu_2) < 0\}$. On $z_0 \in \mathcal{L}_1$, by the monotonicity of $S : x \mapsto e^x/(1 + e^x)$,

$$\begin{aligned} p(\alpha_0, z_0) &= \iint S\big(z_0^\top H(z + \mu_2) + z_0^\top H(\mu_1 - \mu_2) + \alpha_0 + \alpha\big) dF_\alpha(\alpha) dF(z) \\ &> \iint S\big(z_0^\top H(z + \mu_2) + \alpha_0 + \alpha\big) dF_\alpha(\alpha) dF(z) = q(\alpha_0, z_0), \end{aligned}$$

where we use $F_1(z) = F(z - \mu_1)$ and $F_2(z) = F(z - \mu_2)$. By monotonicity of the mapping $x \mapsto x/(1-x)$ for $x \in (0, 1)$, $p/(1-p) > q/(1-q)$ on $\mathcal{L}_1$. We obtain $\big[\{p/(1-p)\}/\{q/(1-q)\}\big]^{A_0^{(1)} - A_0^{(2)}} > 1$, whence we conclude that $G(z_0) > 0$ for $z_0 \in \mathcal{L}_1$. Finally, we have

$$\begin{aligned} I_1 - I_2 &= \int_{\mathcal{L}_1} G(z_0) dF_1(z_0) + \int_{\mathcal{L}_2} G(z_0) dF_1(z_0) \\ &= \int_{\mathcal{L}_1} G(z_0) dF_1(z_0) - \int_{\mathcal{L}_1} G(z_0) dF_2(z_0) \\ &= \int_{\mathcal{L}_1} G(z_0) \big\{f_1(z_0) - f_2(z_0)\big\} dz_0 > 0. \end{aligned}$$

The first equality holds by the assumption that $\mathcal{L}_1 = \{z : z^\top(\mu_1 - \mu_2) > 0\}$ and $\{z^m : z \in \mathcal{L}_1\} = \mathcal{L}_2$. The last inequality holds as $f_1(z_0) > f_2(z_0)$ on $z_0 \in \mathcal{L}_1$ by assumption. The proof is complete after applying the same argument to the case $A_0^{(1)} < A_0^{(2)}$, which implies $I_1 < I_2$. ■

Remark 3 *If $\mu_1 - \mu_2$ is an eigenvector of H associated with a positive eigenvalue λ as in Assumption 4, then the two hyperplanes $\{z : z^\top H(\mu_1 - \mu_2) = 0\}$ and $\{z : z^\top(\mu_1 - \mu_2) = 0\}$ coincide, and for all z such that $z^\top(\mu_1 - \mu_2) > 0$, $z^\top H(\mu_1 - \mu_2) = \lambda z^\top(\mu_1 - \mu_2) > 0$. Furthermore, defining $v = (\mu_1 - \mu_2)/\|\mu_1 - \mu_2\|_2$ as in the proof of Lemma 2, we have*

$$(z_1^m)^\top H z_2^m = z_1^\top (I - 2vv^\top) H (I - 2vv^\top) z_2 = z_1^\top H z_2.$$

Remark 4 *If we can write the density as $f(z) = r(\|z\|_2)$ for some monotone decreasing function $r : \mathbb{R}_+ \to \mathbb{R}_+$, the conditions on the density in Lemma 2 are satisfied.*

In light of the above remarks, we arrive at the fundamental testing lemma for our setup. We only need α's being i.i.d. for Lemma 5 to hold; here the distributional restrictions of α in (8)–(12) of Assumption 1 are superfluous.

Lemma 5 *Consider the testing problem in* (13) *with F being $N_d(0, \tau^2 I_d)$ and $\mu_1 = \mu$ and $\mu_2 = -\mu$. Suppose that Assumptions 1 and 4 hold. Then the likelihood ratio test for the above hypothesis testing problem* (13) *is equivalent to the simple edge counting test where we reject H_0 when $A_0^{(1)} < A_0^{(2)}$.*

3.3.2 Error rates for edge counting

We derive the error rates for edge counting. Consider the testing problem (13) with $F = N_d(0, \tau^2 I_d)$, where $F_+ = N_d(\mu, \tau^2 I_d)$ is the latent distribution for the first community and $F_- = N_d(-\mu, \tau^2 I_d)$ for the second. From now on, write $A_{0,+} = A_0^{(1)} = \sum_{i=1}^m A_{0,i}$ and $A_{0,-} = A_0^{(2)} = \sum_{i=m+1}^{2m} A_{0,i}$. Let ν_n be the probability of making type I+II errors of the test that rejects H_0 in (13) when $A_{0,+} < A_{0,-}$. For any fixed α_0 and z_0, let $p(\alpha_0, z_0)$ and $q(\alpha_0, z_0)$ be defined as in (14) and (15) respectively, and let

$$I(\alpha_0, z_0) = -2\log\left(\{p(\alpha_0, z_0)q(\alpha_0, z_0)\}^{1/2} + [\{1 - p(\alpha_0, z_0)\}\{1 - q(\alpha_0, z_0)\}]^{1/2}\right) \tag{16}$$

be the Rényi divergence of order 1/2 between two Bernoulli distributions Bernoulli($p(\alpha_0, z_0)$) and Bernoulli($q(\alpha_0, z_0)$). The projection distance from μ to the hyperplane $\{z : z^\top H\mu = 0\}$ is then

$$\rho = \frac{\mu^\top H \mu}{(\mu^\top H^2 \mu)^{1/2}}. \tag{17}$$

Furthermore, for any positive integer n and any fixed $\epsilon > 0$, define

$$\overline{\nu}_n^\epsilon = \mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1-\epsilon)I(\alpha_0, z_0)\right\}\right] + \exp\left\{-(1-\epsilon)\frac{\rho^2}{2\tau^2}\right\}, \tag{18}$$

$$\underline{\nu}_n^\epsilon = \mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1+\epsilon)I(\alpha_0, z_0)\right\}\right] + \exp\left\{-(1+\epsilon)\frac{\rho^2}{2\tau^2}\right\}, \tag{19}$$

where $\mathcal{B}_\epsilon = \{z_0 : \|z_0 - \mu\|_2 \leqslant (1 - \epsilon/4)^{1/2}\rho\}$ and the notation $\mathbb{E}_{H_0}^{\alpha_0, z_0}$ means taking expectation over α_0 and z_0 when the null hypothesis in (13) is true. We have $\overline{\nu}_n^0 = \underline{\nu}_n^0$ if we generalize both (18) and (19) to allow $\epsilon = 0$. There are two terms in both (18) and (19). The first term involving the Rényi divergence has previously appeared in the blockmodel community detection literature (Zhang and Zhou, 2016; Jog and Loh, 2015). It reflects the average influence on signal-to-noise ratio from the difference in Bernoulli sampling probabilities of edges connecting nodes within the same or between two different communities. Since the Bernoulli sampling probabilities depend on the realized latent positions, the term collects indirect influence on signal-to-noise ratio from the latent space. The second term depends on the distributions of z's and the quadratic form matrix H only, and it sums up the direct influence on signal-to-noise from the latent space.

With the foregoing definitions, the following lemma controls ν_n from both sides.

Lemma 6 *Suppose that Assumptions 1 and 3 hold. Let $n = 2m+1$ and that $z_i \overset{iid}{\sim} N_d(\mu, \tau^2 I_d)$ for $i = 1, \ldots, m$ and $z_i \overset{iid}{\sim} N_d(-\mu, \tau^2 I_d)$ for $i = m+1, \ldots, 2m$, where $\tau \to 0$ as $n \to \infty$. Further, assume that $\overline{\nu}_n^0 \to 0$ as $n \to \infty$, then for any $\epsilon \in (0, 1/2)$, there is an n_ϵ such that for all $n > n_\epsilon$,*

$$\underline{\nu}_n^\epsilon \leqslant \nu_n \leqslant \overline{\nu}_n^\epsilon. \tag{20}$$

The proof of Lemma 6 can be found in Appendix A.

3.4 Rates of convergence

In this subsection, we present rates of convergence on errors of our initial and refined estimators.

Upper bounds The following proposition gives upper bounds for estimators obtained from Algorithm 1.

Proposition 7 *Suppose that Assumptions 1, 2 and 3 hold. Assume that the n nodes have true labels σ, where $\sigma_i = 1$ for $i = 1, \ldots, n_1$, $\sigma_i = 2$ for $i = n_1+1, \ldots, n$, and for $n_2 = n - n_1$, $n_1, n_2 \in \left[(1-\delta_n)n/2, (1+\delta_n)n/2\right]$. Let $\hat{\sigma}^0$ be the output of Algorithm 1. Let ω_i and $\underline{\omega}$ be defined as in Assumption 1. Then for any $\gamma > 0$, some constant $C > 0$ and all sufficiently large n, we have*

$$\mathbb{P}\{\ell(\sigma, \hat{\sigma}^0) \leqslant \gamma\} \geqslant \mathbb{P}\Bigg(\sum_{\{i: \sigma_i \neq \hat{\sigma}_i^0\}} e^{\omega_i} \leqslant e^{-\underline{\omega}} \gamma n \Bigg) \geqslant 1 - n^{-(1+2C)}.$$

We present the proof of Proposition 7 in Appendix B.

The following theorem gives our main upper bounds on the output of Algorithm 3.

Theorem 8 *Let $k = 2$ and $\mathcal{P}_n = \mathcal{P}_n(H, \mu, \tau, F_\alpha)$. Suppose that Assumptions 1, 2 and 3 hold. For any $\epsilon \in (0, 1/2)$, let $\overline{\nu}_n^\epsilon$ be defined as in (18). Suppose $\overline{\nu}_n^0 \to 0$ as $n \to \infty$. Then for any fixed $\epsilon > 0$, the output $\widehat{\sigma}$ of Algorithm 3 satisfies*

$$\limsup_{n \to \infty} \sup_{\mathcal{P}_n} \mathbb{P}\left\{\ell(\sigma, \widehat{\sigma}) > \overline{\nu}_n^\epsilon\right\} = 0.$$

The high probability upper bound in Theorem 8 consists of two terms as on the right hand side of (18). In view of the discussion following (18), the first term summarizes influence on the clustering error from the network signal, averaged over realizations of degree sequence and latent positions. Hence we regard it as the network term. The second term collects immediate influence on clustering error by signal from latent space as it depends only on H and the latent position distributions, which could be viewed as the latent space term.

Lower bounds We conclude this section with the following minimax lower bounds when Assumption 4 holds, which implies Assumption 3. The lower bounds match the upper bounds in Theorem 8 up to some arbitrarily small perturbation of the exponents.

Theorem 9 *Let $k = 2$ and $\mathcal{P}_n = \mathcal{P}_n(H, \mu, \tau, F_\alpha)$. Suppose that Assumptions 1, 2 and 4 hold. Suppose $\overline{\nu}_n^0 \to 0$ as $n \to \infty$. For any $\epsilon \in (0, 1/2)$, define $\underline{\nu}_n^\epsilon$ as in (19), then the minimax risk satisfies*

$$\inf_{\widehat{\sigma}} \sup_{\mathcal{P}_n} \mathbb{E}\{\ell(\sigma, \widehat{\sigma})\} \gtrsim \underline{\nu}_n^\epsilon. \tag{21}$$

The proofs of Theorems 8 and 9 are given in Appendix C.

Remark 10 *In view of the discussion following (4), in the anisotropic case where $F_{z,j} = N_d((-1)^{j-1}\mu, \tau^2\Sigma)$, Theorems 8 holds with ρ redefined by $\rho = (\mu^\top H\mu)/(\mu^\top H\Sigma H\mu)^{1/2}$ and $\mathcal{B}_\epsilon$ redefined by $\mathcal{B}_\epsilon = \{z_0 : \|\Sigma^{-1/2}(z_0 - \mu)\|_2 \leqslant (1 - \epsilon/4)^{1/2}\rho\}$ in (18) and (19). Theorem 9 holds with the same redefinitions of ρ and $\mathcal{B}_\epsilon$, and Assumption 4 replaced by that μ is an eigenvector of ΣH associated with some positive eigenvalue.*

4. Simulation studies

In this section, we evaluate numerical performance of both SpecLoRe and Algorithm 3 on simulated examples generated according to four different parameter specifications of the latent space model. All reported results were obtained on a Windows 7 PC with two Intel Xeon Processors (E5-2630 v3@2.40GHz) and 64G RAM.

Specification 1 We first consider the case where H is positive semi-definite. In this case, we compare both SpecLoRe and Algorithm 3 with the LSCD method.

We set up model (1) with latent space dimension $d = 3$ and size $n = 1000$. The nodes were split into two clusters of sizes $n_1 = n_2 = 500$. For $i = 1, \ldots, n_1$, we generated i.i.d. $z_i \sim N_d(\mu, \tau^2 I_d)$, where $\mu = (0.5, 1, 0)^\top$, and for $i = n_1 + 1, \ldots, n$, we generated i.i.d. $z_i \sim N_d(-\mu, \tau^2 I_d)$. We varied $\tau \in \{0.75, 0.5, 0.25\}$. In addition, we let $H = \mathrm{diag}(1, 1, 0.5)$, and generated $\alpha_i = \overline{\alpha} + \omega_i$, where $\overline{\alpha} = -2.49$ (so that the median degree $ne^{2\overline{\alpha}} = \log n$) and $\omega_i \overset{iid}{\sim} N(0, 1)$. We have designed the setting so that μ is an eigenvector of H with positive eigenvalue 1. In each repetition, we generated one copy of the adjacency matrix A with diagonals $A_{ii} = 0$ for $i \in [n]$. Then we applied the SpecLoRe method with $R = 1$ and $R = 10$ rounds of local refinement to cluster nodes. We also ran Algorithm 3 to investigate its numerical difference from SpecLoRe. For LSCD, we used Algorithm 3 in Ma et al. (2020) as the initializer, then applied Algorithm 1 in Ma et al. (2020) with 800 iterations followed by k-means clustering.

Table 1 reports average misclustering proportions (7) over 100 repetitions and average runtimes (in seconds) of SpecLoRe (denoted "SpecLoRe" with subscripts $R = 1$ and

$R = 10$), Algorithm 3 and LSCD. The runtime of SpecLoRe included time spent on spectral initialization by Algorithm 1. It also reports average degrees (namely the average of $(1/n)\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}$ over 100 repetitions). Furthermore, it reports theoretical Bayes risks, which are the best possible misclustering errors if we observe the latent positions directly and know the underlying distributions that generated the z_i's. Bayes risk is only attainable by reconstructing the underlying distributions based on infinite samples directly observed from the latent variable distributions. Finally, the "Initial" column reports the average errors of the initial estimates obtained from Algorithm 1.

τ	Avg degree	Bayes risk	LSCD		Algo3	Initial	$\text{SpecLoRe}_{R=1}$		$\text{SpecLoRe}_{R=10}$	
			error	time	error	error	error	time	error	time
0.75	47.68	6.80%	8.03%	179.29	8.27%	8.33%	8.21%	2.10	8.20%	2.72
0.5	35.28	1.27%	2.93%	184.31	3.20%	3.44%	3.18%	2.07	3.18%	2.63
0.25	29.51	3.87E-4%	0.82%	182.72	0.84%	1.36%	0.85%	2.02	0.83%	2.63

Table 1: Misclustering proportions and runtimes in Specification 1.

For all three values of τ, misclustering errors of SpecLoRe with $R = 10$ and LSCD were close, but runtimes of the former method were only tiny proportions of those of the latter. We also observe that misclustering errors of SpecLoRe with $R = 1$ were nearly identical to those of Algorithm 3. This reassures that repeated initializations in Algorithm 3 were only needed for technical reasons in proofs, and justifies the use of SpecLoRe in practice. Furthermore, for $\tau = 0.75$, the misclustering errors of SpecLoRe were close to Bayes risk, while for $\tau = 0.25$ the misclustering errors of SpecLoRe were much larger than Bayes risk. This suggests that when τ is large, the signal-to-noise ratio affected by the latent positions dominates the error rate, while when τ is small, the signal-to-noise ratio affected by the network sparsity dominates.

Specification 2 In the second study, we kept the same settings as in the first case except that we set $H = \text{diag}(1, 1, -0.5)$ which is no longer positive semi-definite, while μ is still an eigenvector of H with eigenvalue 1. In this case, the LSCD method cannot be directly applied, and so we did not report its results in this case. Table 2 reports all the other columns in Table 1 in the present setting. Overall, misclustering errors and runtimes of various algorithms in this setting were almost identical to those in the first study.

τ	Avg degree	Bayes risk	Algo3	Initial	$\text{SpecLoRe}_{R=1}$		$\text{SpecLoRe}_{R=10}$	
			error	error	error	time	error	time
0.75	47.85	6.80%	8.25%	8.28%	8.18%	2.13	8.16%	2.68
0.5	35.41	1.27%	3.16%	3.44%	3.16%	2.18	3.14%	2.73
0.25	29.51	3.87E-4%	0.82%	1.31%	0.85%	2.12	0.79%	2.65

Table 2: Misclustering proportions and runtimes in Specification 2.

Specification 3 In the third study, the settings remained the same as in the first study except that we fixed $\tau = 0.5$ and let $\overline{\alpha} \in \{-2.14, -2.49, -2.83\}$, which calibrated the median degree of networks to be around $\{2, 1, 0.5\} \times \log n$, respectively. Table 3 reports the results for all three different $\overline{\alpha}$'s. As $|\overline{\alpha}|$ grows, the average degree decreases significantly. Misclustering errors of SpecLoRe with $R = 10$ were slightly worse than those of the LSCD method, but were always within 110% of the LSCD errors. On the other hand, runtimes of SpecLoRe with $R = 10$ were of smaller order of magnitude than those of LSCD. Misclustering errors of SpecLoRe were comparable to Bayes risk when $\overline{\alpha} = -2.14$, and became more sizeable relative to Bayes risk for larger $\overline{\alpha}$. This suggests that network sparsity becomes the dominating factor in error rate as $|\overline{\alpha}|$ grows.

$\overline{\alpha}$	Avg degree	Bayes risk	LSCD		Algo3	Initial	SpecLoRe$_{R=1}$		SpecLoRe$_{R=10}$	
			error	time	error	error	error	time	error	time
-2.14	58.86	1.27%	2.04%	219.92	2.24%	2.27%	2.25%	2.04	2.23%	2.59
-2.49	35.28	1.27%	2.93%	211.29	3.20%	3.44%	3.18%	2.31	3.17%	2.86
-2.83	20.30	1.27%	4.58%	213.31	4.94%	6.04%	4.91%	2.26	4.88%	2.85

Table 3: Misclustering proportions and runtimes in Specification 3.

Specification 4 Finally, we repeated the last two studies with $H = \text{diag}(1, 1, -0.5)$ and $\mu = (1.25/1.29)^{1/2}\,(0.5, 1, 0.2)^\top$. In this case, μ is no longer an eigenvector of H but $\|\mu\|_2$ is the same as in specifications 1–3 to make the results more comparable. Table 4 summarizes the relevant results for all different combinations of τ and $\overline{\alpha}$ values. We observe that the first three rows had slightly larger misclustering errors than those in Tables 1 and 2, and the last three rows had slightly larger misclustering errors than those in Table 3. Such a difference conforms with our theory since quantity ρ (defined in (17)) in (18)–(19) becomes smaller when μ is no longer an eigenvector of H with maximum possible eigenvalue 1 under (5), resulting in larger error rates.

τ	$\overline{\alpha}$	Avg degree	Bayes risk	Algo3	Initial	SpecLoRe$_{R=1}$		SpecLoRe$_{R=10}$	
				error	error	error	time	error	time
0.75	-2.49	46.34	6.80%	8.89%	8.89%	8.83%	2.27	8.80%	2.82
0.5	-2.49	34.09	1.27%	3.63%	3.93%	3.62%	2.16	3.62%	2.71
0.25	-2.49	28.55	3.87E-4%	0.97%	1.56%	1.01%	2.11	1.00%	2.68
0.5	-2.14	57.64	1.27%	2.55%	2.60%	2.53%	2.07	2.53%	2.63
0.5	-2.49	34.09	1.27%	3.51%	3.93%	3.62%	2.16	3.62%	2.71
0.5	-2.83	19.72	1.27%	5.35%	6.45%	5.33%	2.15	5.27%	2.73

Table 4: Misclustering proportions and runtimes in Specification 4.

5. Real data examples

We now demonstrate performance of the proposed algorithm on some real data examples. More detailed comparison of Algorithm 3 with Algorithms 1+2 and other methods on carefully constructed simulated examples can be found in Section 4 of the appendices.

We consider five datasets. The first three datasets are political blog with 1222 nodes, 16714 edges, and 2 communities (Adamic and Glance, 2005), Simmons College with 1137 nodes, 24257 edges, and 4 communities and Caltech data with 590 nodes, 12822 edges, and 8 communities (Traud et al., 2011, 2012). For Simmons College and Caltech data, we followed the same pre-processing steps as in Chen et al. (2018). These datasets have been studied extensively in the blockmodel community detection literature.

The fourth dataset is a manufacturing company network from Cross and Parker (2004), which was studied in Weng and Feng (2022). Questions were asked to pairs of employees on their ties in work, and weights were assigned on a 0–6 scale where higher weights correspond to closer ties. Following Weng and Feng (2022), we used the weights to create an adjacency matrix: We set $A_{ij} = A_{ji} = 1$ if and only if both edges from i to j and from j to i have weights larger than 3. Otherwise, $A_{ij} = A_{ji} = 0$. This resulted in an undirected network with 74 nodes and 235 edges. Four communities were formed according to the "location" value of each node which is the most assortative among three available nodes attributes in this data.

The fifth dataset is a French high school friendship network (Mastrandrea et al., 2015). This dataset recorded friendship relations and contacts among 329 students in a Marseilles high school. To construct an adjacency matrix, we took the first contact information which recorded active contacts between students during 20-second intervals of the data collection process over a measuring infrastructure. We set $A_{ij} = A_{ji} = 1$ if and only if there were contacts recorded between i and j. The resulting network has 5818 edges. Each student belonged to one of nine classes which we regarded as nine true communities.

We compare Algorithm 1 + one-round Algorithm 2 refinement (SpecLoRe$_{R=1}$) and Algorithm 1 + ten-round Algorithm 2 refinement (SpecLoRe$_{R=10}$) to LSCD in Ma et al. (2020) (initialized by Algorithm 3 in Ma et al., 2020 followed by Algorithm 1 in Ma et al., 2020 with 800 iterations). Algorithm 3 has essentially the same level of accuracy as SpecLoRe with $R = 1$, which we have illustrated in detail in Section 4. The LSCD methods functioned as the benchmark. Comparison of LSCD to several other state-of-the-art methods (SCORE (Jin, 2015), OCCAM (Zhang et al.), and CMM (Chen et al., 2018)) on the first three datasets was conducted in Ma et al. (2020). LSCD was shown to be a top performer, and so we omit comparison to other methods on the first three datasets. We set latent space dimension equal to number of communities for LSCD.

		LSCD		Initial	SpecLoRe$_{R=1}$		SpecLoRe$_{R=10}$	
Dataset	# Clusters	error	time	error	error	time	error	time
Political blog	2	4.91%	43.31	5.32%	4.66%	0.62	4.66%	0.97
Simmons	4	11.87%	39.90	13.54%	11.61%	1.94	11.17%	2.65
Caltech	8	18.14%	11.85	21.69%	17.46%	0.87	14.58%	1.29
Company	4	1.35%	0.83	5.41%	2.70%	0.01	1.35%	0.02
High school	9	0.61%	5.29	0.61%	0.61%	0.13	0.61%	0.24

Table 5: A summary of performances on five datasets. Each "error" column reports proportions of misclustered nodes. Each "time" column reports runtime of the corresponding method in seconds (including initialization).

Table 5 presents performances of both versions of SpecLoRe and those of LSCD in terms of accuracy and speed. For reported speed of SpecLoRe, we have included time spent

	SCORE		OCCAM		CMM	
Dataset	error	time	error	time	error	time
Company	8.11%	0.27	1.35%	0.84	2.70%	0.28
High school	0.61%	0.62	0.61%	4.90	1.82%	2.98

Table 6: A summary of performances of three other community detection methods (SCORE, OCCAM, and CMM) on manufacturing company and French high school datasets. Each "error" column reports proportions of misclustered nodes. Each "time" column reports runtime of the corresponding method in seconds.

on spectral initialization. In addition, it also reports accuracy of spectral initialization (Algorithm 1). On these five datasets, $\text{SpecLoRe}_{R=10}$ and LSCD were comparable in terms of accuracy while $\text{SpecLoRe}_{R=10}$ was significantly faster (and slightly more accurate in most examples). This is not surprising because it aims only at clustering nodes while LSCD fits all parameters. $\text{SpecLoRe}_{R=1}$ was the fastest due to a single round of refinement which incurred the cost of slightly inferior accuracy. However, it still notably improved the accuracy of spectral clustering. For benchmarking purpose, the performances of three competitive community detection methods, namely, SCORE, OCCAM, and CMM, on the fourth (Company) and the fifth (High school) datasets are reported in Table 6. Compared with the last two rows in Table 5, $\text{SpecLoRe}_{R=10}$ continues to be the best when both accuracy and speed are taken into account. All reported results were obtained on a Windows 7 PC with two Intel Xeon Processors (E5-2630 v3@2.40GHz) and 64G RAM.

6. Discussions

In this paper, we study theoretical and empirical performances of a simple community detection algorithm in the context of sparse latent space models. We establish consistency and rates of convergence of the method for sparse latent eigenmodels with two balanced communities. Under an additional eigenvector assumption (Assumption 4), we further argue that our rate has sharp exponent in a minimax sense. Although we have centered our theoretical investigations on balanced two community case, the method performs well empirically in more general scenarios encountered in real world data examples. Under current setup, an immediate future research direction is to see whether the same upper bound can be established for Algorithms 1 and 2 directly.

It is natural to extend the current theoretical framework to cases where $k > 2$, all communities have roughly equal sizes, and each component of the latent mixture distribution is sub-Gaussian and isotropic. We expect an analogous error rate of our proposed algorithm to hold with a possibly gruesome but direct analysis by generalizing Lemma 6 to the case $k > 2$, and then subsequently Theorem 8. If $\|\mu_i\|_2$ for all $i \in [k]$ are all the same and $\mu_i - \mu_j$ for $1 \leqslant i < j \leqslant k$ are all eigenvectors of H associated with positive eigenvalues, we may employ the key Lemma 2 to carry out pairwise analysis for each community pair (j_1, j_2) $(1 \leqslant j_1 < j_2 \leqslant k)$, which gives us the equivalence between the optimal (pairwise) likelihood ratio tests and edge counting. This would pave the way for matching lower bound by a generalized version of Theorem 9.

A more challenging future research direction is to generalize the current framework to handle non-homogeneous mixture distributions of latent variables. For instance, if we assume that the latent variable $z \sim N_d(\mu, \Sigma_1)$ when the node is in community 1 and $z \sim N_d(-\mu, \Sigma_2)$ in community 2 with $\Sigma_1 \neq \Sigma_2$, the problem becomes more difficult where new understandings and techniques need to be discovered. First, the upper bound analysis will be more entangled after losing homogeneity (and isotropy) as our analysis exploits various symmetries whenever possible. Moreover, it is even less clear whether it is possible to establish something akin to Lemma 2, which bridges the edge-counting procedure and the optimal likelihood ratio test so that a matching lower bound would be in sight. The reason is that the proof of the current Lemma 2 relies crucially on exploiting subtle symmetric structures, which is no longer true when the latent space is distorted by the non-homogeneity.

We have focused on the case where one only observes a network structure among n nodes. An important advantage of latent space models is the convenience to further include node and/or edge covariates (Hoff et al., 2002). Though it is beyond the scope of the present paper, it is nonetheless desirable to understand how the presence of covariates could affect community detection on nodes. Furthermore, whether there is covariate or not, it is of interest to explore information-theoretic limits and optimal algorithms for community detection when Assumption 4 fails.

Acknowledgements

FG's research was supported in part by NSFC grants 11701095 and 11690013.

Appendix A. Proof of Lemma 6

By Jensen's inequality, for any fixed $\epsilon \in (0, 1/2)$,

$$\overline{\nu}_n^{\epsilon} \leqslant (\overline{\nu}_n^{0})^{1-\epsilon} \to 0,$$

as $n \to \infty$. By symmetry, we have

$$\begin{aligned}\nu_n &= \mathbb{P}_{H_0}(A_{0,+} < A_{0,-}) + \mathbb{P}_{H_1}(A_{0,+} \geqslant A_{0,-}) \\ &= \mathbb{P}_{H_0}(A_{0,+} < A_{0,-}) + \mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}).\end{aligned}$$

Hence,

$$\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}) \leqslant \nu_n \leqslant 2\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}). \tag{22}$$

Upper bound By law of total expectation,

$$\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}) = \mathbb{E}_{H_0}^{\alpha_0, z_0}\big\{\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0)\big\}.$$

Let

$$\Omega = \left\{ \{a_{0,i}\}_{i=1}^{2m} : a_{0,i} \in \{0,1\} \text{ for } 1 \leqslant i \leqslant 2m, \sum_{i=1}^{m} a_{0,i} \leqslant \sum_{i=m+1}^{2m} a_{0,i} \right\}.$$

We then have

$$
\begin{aligned}
&\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0) \\
&= \sum_{\{a_{0,i}\}_{i=1}^{2m} \in \Omega} \mathbb{P}(A_{0,1} = a_{0,1}, \ldots, A_{0,2m} = a_{0,2m} \mid \alpha_0, z_0) \\
&= \sum_{\{a_{0,i}\}_{i=1}^{2m} \in \Omega} \mathbb{E}^{\{\alpha_i, z_i\}_{i=1}^{2m}} \big\{ \mathbb{P}(A_{0,1} = a_{0,1}, \ldots, A_{0,2m} = a_{0,2m} \mid \alpha_0, z_0, \{\alpha_i, z_i\}_{i=1}^{2m}) \big\} \\
&= \sum_{\{a_{0,i}\}_{i=1}^{2m} \in \Omega} \mathbb{E}^{\{\alpha_i, z_i\}_{i=1}^{2m}} \left\{ \prod_{i=1}^{2m} \mathbb{P}(A_{0,i} = a_{0,i} \mid \alpha_0, z_0, \alpha_i, z_i) \right\} \\
&= \sum_{\{a_{0,i}\}_{i=1}^{2m} \in \Omega} \prod_{i=1}^{2m} \mathbb{E}^{\alpha_i, z_i} \big\{ \mathbb{P}(A_{0,i} = a_{0,i} \mid \alpha_0, z_0, \alpha_i, z_i) \big\}.
\end{aligned}
$$

Here $\mathbb{E}^{\alpha_i, z_i}$ means the expectation over α_i and z_i (under H_0). In the last equality, we have used the mutual independence of $\{\alpha_i, z_i\}$ for $1 \leqslant i \leqslant 2m$. By the discussion preceding (13) and the definition in (14) and (15), we have

$$
\mathbb{E}^{\alpha_i, z_i} \big\{ \mathbb{P}(A_{0,i} = 1 \mid \alpha_0, z_0, \alpha_i, z_i) \big\} = \begin{cases} p(\alpha_0, z_0), & 1 \leqslant i \leqslant m, \\ q(\alpha_0, z_0), & m+1 \leqslant i \leqslant 2m. \end{cases}
$$

By definition, $p(\alpha_0, z_0)$ and $q(\alpha_0, z_0)$ can be written as

$$
p(\alpha_0, z_0) = \mathbb{E}^{\alpha_1, z_1} S(z_0^\top H z_1 + \alpha_0 + \alpha_1), \tag{23}
$$

$$
\begin{aligned}
q(\alpha_0, z_0) &= \mathbb{E}^{\alpha_{m+1}, z_{m+1}} S(z_0^\top H z_{m+1} + \alpha_0 + \alpha_{m+1}) \\
&= \mathbb{E}^{\alpha_1, z_1} S(-z_0^\top H z_1 + \alpha_0 + \alpha_1).
\end{aligned} \tag{24}
$$

Here $\alpha_i \overset{iid}{\sim} F_\alpha$, $z_1 \sim N(\mu, \tau^2 I_d)$ and $z_{m+1} \sim N(-\mu, \tau^2 I_d)$, and they are mutually independent. Define $\mathcal{L}_+ = \{z_0 : z_0^\top H\mu \geqslant 0\}$ and $\mathcal{L}_- = \{z_0 : z_0^\top H\mu < 0\}$. Conditional on α_0 and z_0, the distribution of $z_0^\top H(z_1 - \mu)$ is symmetric about zero and is independent of α_1. Since S is a monotone increasing function, together with (23) and (24), this observation implies that $p(\alpha_0, z_0) \geqslant q(\alpha_0, z_0)$ when $z_0 \in \mathcal{L}_+$ and $p(\alpha_0, z_0) < q(\alpha_0, z_0)$ when $z_0 \in \mathcal{L}_-$.

For any $z_0 \in \mathcal{B}_\epsilon$, we have

$$
\begin{aligned}
z_0^\top H\mu &= \mu^\top H\mu + (z_0 - \mu)^\top H\mu \\
&\geqslant \mu^\top H\mu - |(z_0 - \mu)^\top H\mu| \\
&\geqslant \mu^\top H\mu - \|H\mu\|_2 \|z_0 - \mu\|_2 \\
&\geqslant \mu^\top H\mu - (\mu^\top H^2 \mu)^{1/2} (1 - \epsilon/4)^{1/2} \rho \\
&= \left\{ 1 - (1 - \epsilon/4)^{1/2} \right\} \mu^\top H\mu \\
&\geqslant \epsilon \mu^\top H\mu / 8.
\end{aligned} \tag{25}
$$

Here the second equality holds due to (17). Thus, $\mathcal{B}_\epsilon \subset \mathcal{L}_+$. See Figure 1 for a graphical illustration.

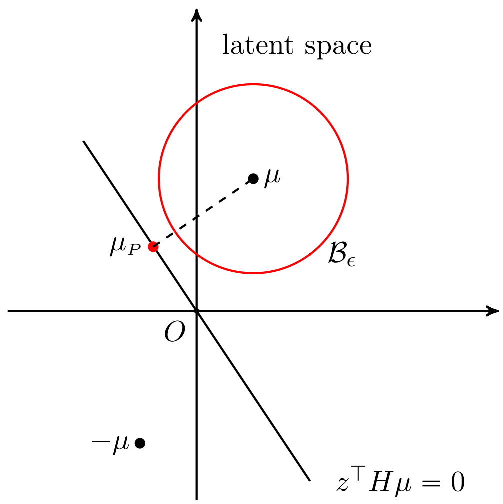

Figure 1: An illustration of a $\mathcal{B}_\epsilon$-ball in the latent space: μ_P is the orthogonal projection of μ onto the hyperplane $\{z : z^\top H\mu = 0\}$ with the distance between μ and μ_P equal to ρ defined in (17). Given $\epsilon > 0$, $\mathcal{B}_\epsilon$ is the ball in red with radius $(1 - \epsilon/4)^{1/2}\rho$.

Next, we derive uniform bounds of $p(\alpha_0, z_0)$, $q(\alpha_0, z_0)$ and $I(\alpha_0, z_0)$ for all $z_0 \in \mathcal{B}_\epsilon$. To this end, define

$$D_p(\omega_0, z_0) = \mathbb{E}^{\omega_1, z_1}\left\{\frac{e^{z_0^\top H(z_1-\mu)+\omega_0+\omega_1}}{1+e^{z_0^\top Hz_1+2\overline{\alpha}+\omega_0+\omega_1}}\right\}, \quad D_q(\omega_0, z_0) = \mathbb{E}^{\omega_1, z_1}\left\{\frac{e^{-z_0^\top H(z_1-\mu)+\omega_0+\omega_1}}{1+e^{-z_0^\top Hz_1+2\overline{\alpha}+\omega_0+\omega_1}}\right\}.$$

By (23) and (24), we have

$$p(\alpha_0, z_0) = e^{2\overline{\alpha}} e^{z_0^\top H\mu} D_p(\omega_0, z_0) \tag{26}$$

$$q(\alpha_0, z_0) = e^{2\overline{\alpha}} e^{-z_0^\top H\mu} D_q(\omega_0, z_0). \tag{27}$$

To find upper bounds for $D_p(\omega_0, z_0)$ and $D_q(\omega_0, z_0)$, we define

$$D(\omega_0, z_0) = \mathbb{E}^{\omega_1, z_1}\left\{e^{z_0^\top H(z_1-\mu)+\omega_0+\omega_1}\right\} = e^{\omega_0}\mathbb{E}(e^{\omega_1})\mathbb{E}^{z_1}\left\{e^{z_0^\top H(z_1-\mu)}\right\}.$$

Then we have

$$D_p(\omega_0, z_0) \leqslant \mathbb{E}^{\omega_1, z_1}\left\{e^{z_0^\top H(z_1-\mu)+\omega_0+\omega_1}\right\} = D(\omega_0, z_0) \tag{28}$$

$$D_q(\omega_0, z_0) \leqslant \mathbb{E}^{\omega_1, z_1}\left\{e^{-z_0^\top H(z_1-\mu)+\omega_0+\omega_1}\right\} = D(\omega_0, z_0). \tag{29}$$

where the last equality holds since the distribution of $z_1 - \mu$ is symmetric about zero. By Assumption 1, $\mathbb{E}(e^{\omega_1}) \leqslant \left\{\mathbb{E}(e^{2\omega_1})\right\}^{1/2} \leqslant C^{1/2}$. This inequality, combined with the boundedness of z_0 for $z_0 \in \mathcal{B}_\epsilon$ and (8) of Assumption 1 implies that

$$0 < e^{-2\underline{\omega}}\underline{D} \leqslant D(\omega_0, z_0) \leqslant e^{\omega'}\overline{D}, \tag{30}$$

where $\overline{D}$ and $\underline{D}$ are constants.

On the other hand, to find lower bounds for $D_p(z_0, \omega_0)$ and $D_q(z_0, \omega_0)$, we define

$$D_2(\omega_0, z_0) = \mathbb{E}^{\omega_1, z_1}\left\{e^{2z_0^\top H(z_1-\mu)+2\omega_0+2\omega_1}\right\} = e^{2\omega_0}\mathbb{E}(e^{2\omega_1})\mathbb{E}^{z_1}\left\{e^{2z_0^\top H(z_1-\mu)}\right\}.$$

By Assumption 1, $\mathbb{E}(e^{2\omega_1}) \leqslant C$. Further by (8) of Assumption 1 and boundedness of z_0, $D_2(\omega_0, z_0)$ also has an upper bound $e^{2\omega'}\overline{D}_2$ where $\overline{D}_2$ is a constant. Then

$$\begin{aligned}
D(\omega_0, z_0) - D_p(\omega_0, z_0) =& \mathbb{E}^{\omega_1, z_1}\left\{e^{z_0^\top H(z_1-\mu)+\omega_0+\omega_1}\left(1 - \frac{1}{1+e^{z_0^\top Hz_1+2\overline{\alpha}+\omega_0+\omega_1}}\right)\right\} \\
=& e^{2\overline{\alpha}}e^{z_0^\top H\mu}\mathbb{E}^{\omega_1, z_1}\left\{\frac{e^{2z_0^\top H(z_1-\mu)+2\omega_0+2\omega_1}}{1+e^{z_0^\top Hz_1+2\overline{\alpha}+\omega_0+\omega_1}}\right\} \\
\leqslant& e^{2\overline{\alpha}}e^{z_0^\top H\mu}\mathbb{E}^{\omega_1, z_1}\left\{e^{2z_0^\top H(z_1-\mu)+2\omega_0+2\omega_1}\right\} \\
=& e^{2\overline{\alpha}}e^{z_0^\top H\mu}D_2(\omega_0, z_0) \\
\leqslant& e^{2\overline{\alpha}+2\omega'}e^{z_0^\top H\mu}\overline{D}_2.
\end{aligned} \tag{31}$$

Let $0 < \kappa < 1$ be any fixed constant. By (9) of Assumption 1 and the boundedness of z_0 within $\mathcal{B}_\epsilon$, the inequality $e^{2\overline{\alpha}+2\omega'}\exp(z_0^\top H\mu)\overline{D}_2 \leqslant \kappa e^{-2\underline{\omega}}\underline{D}$ holds for all sufficiently large n. By (30),

$$e^{2\overline{\alpha}+2\omega'}e^{z_0^\top H\mu}\overline{D}_2 \leqslant \kappa e^{-2\underline{\omega}}\underline{D} \leqslant \kappa D(\omega_0, z_0). \tag{32}$$

Combining (31) and (32), we have

$$D_p(\omega_0, z_0) \geqslant (1-\kappa)D(\omega_0, z_0). \tag{33}$$

By the same argument, we can also get

$$D_q(\omega_0, z_0) \geqslant (1-\kappa)D(\omega_0, z_0). \tag{34}$$

We now derive a lower bound for $I(\alpha_0, z_0)$. By definition, we have

$$\begin{aligned}
I(\alpha_0, z_0) =& -2\log\left(\{p(\alpha_0, z_0)q(\alpha_0, z_0)\}^{1/2} + [\{1-p(\alpha_0, z_0)\}\{1-q(\alpha_0, z_0)\}]^{1/2}\right) \\
\geqslant& -2\log\left[\{p(\alpha_0, z_0)q(\alpha_0, z_0)\}^{1/2} + 1 - \frac{1}{2}\{p(\alpha_0, z_0)+q(\alpha_0, z_0)\}\right] \\
\geqslant& -2\{p(\alpha_0, z_0)q(\alpha_0, z_0)\}^{1/2} + p(\alpha_0, z_0) + q(\alpha_0, z_0) \\
=& e^{2\overline{\alpha}}e^{z_0^\top H\mu}\left[\{D_p(\omega_0, z_0)\}^{1/2} - e^{-z_0^\top H\mu}\{D_q(\omega_0, z_0)\}^{1/2}\right]^2,
\end{aligned}$$

where the last inequality is due to $\log(1-x) \leqslant -x$ for $0 < x < 1$. We let

$$C(\omega_0, z_0) = e^{z_0^\top H\mu}\left[\{D_p(\omega_0, z_0)\}^{1/2} - e^{-z_0^\top H\mu}\{D_q(\omega_0, z_0)\}^{1/2}\right]^2,$$

and let $\kappa = 1 - \{1 + \exp(-\epsilon\mu^\top H\mu/8)\}^2/4$. Then by (25), (29) and (33) we get

$$\begin{aligned}
C(\omega_0, z_0) \geqslant & e^{\frac{\epsilon}{8}\mu^\top H\mu}\left[\{(1-\kappa)D(\omega_0, z_0)\}^{1/2} - e^{-z_0^\top H\mu}\{D(\omega_0, z_0)\}^{1/2}\right]^2 \\
= & e^{\frac{\epsilon}{8}\mu^\top H\mu} D(\omega_0, z_0)\left\{(1-\kappa)^{1/2} - e^{-z_0^\top H\mu}\right\}^2 \\
\geqslant & e^{\frac{\epsilon}{8}\mu^\top H\mu} D(\omega_0, z_0)\left\{\frac{1}{2}\left(1 + e^{-\frac{\epsilon}{8}\mu^\top H\mu}\right) - e^{-z_0^\top H\mu}\right\}^2 \\
\geqslant & e^{\frac{\epsilon}{8}\mu^\top H\mu} D(\omega_0, z_0)\left\{\frac{1}{2}\left(1 + e^{-\frac{\epsilon}{8}\mu^\top H\mu}\right) - e^{-\frac{\epsilon}{8}\mu^\top H\mu}\right\}^2 \\
= & e^{\frac{\epsilon}{8}\mu^\top H\mu} D(\omega_0, z_0)\left\{\frac{1}{2}\left(1 - e^{-\frac{\epsilon}{8}\mu^\top H\mu}\right)\right\}^2 \\
\geqslant & e^{\frac{\epsilon}{8}\mu^\top H\mu} e^{-2\underline{\omega}}\underline{D}\left\{\frac{1}{2}\left(1 - e^{-\frac{\epsilon}{8}\mu^\top H\mu}\right)\right\}^2 .
\end{aligned}$$

We denote the right-hand side of the last inequality as $\underline{C}$. Since $\underline{D}$ and $\underline{\omega}$ are both constants, $\underline{C} > 0$ is also a constant. In summary, for $z_0 \in \mathcal{B}_\epsilon$, we have established

$$I(\alpha_0, z_0) \geqslant e^{2\overline{\alpha}}\underline{C}, \tag{35}$$

where $\underline{C}$ is some constant depending on ϵ.

In view of the foregoing discussion, we can write

$$\begin{aligned}
\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}) = & \mathbb{E}_{H_0}^{\alpha_0, z_0}\left\{1(z_0 \in \mathcal{B}_\epsilon)\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0)\right\} \\
& + \mathbb{E}_{H_0}^{\alpha_0, z_0}\left\{1(z_0 \in \mathcal{B}_\epsilon^c)\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0)\right\}.
\end{aligned} \tag{36}$$

Conditional on α_0 and z_0, we can generate independent random variables $W_i \sim \text{Bernoulli}(p(\alpha_0, z_0))$ for $i = 1, \ldots, m$ and $W_i \sim \text{Bernoulli}(q(\alpha_0, z_0))$ for $i = m+1, \ldots, 2m$. Then we have

$$\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0) = \mathbb{P}\left(\sum_{i=1}^m W_i \leqslant \sum_{i=m+1}^{2m} W_i\right).$$

For any α_0 and any $z_0 \in \mathcal{B}_\epsilon$, aside from $p(\alpha_0, z_0) > q(\alpha_0, z_0)$, we can also get from (26), (27), (28), (29), (30), z_0 bounded, and (9) of Assumption 1 that as $n \to \infty$,

$$p(\alpha_0, z_0) \to 0, \quad q(\alpha_0, z_0) \to 0.$$

We then obtain from the calculation in Gao et al. (2017) and Gao and Ma (2020) that

$$\mathbb{P}\left(\sum_{i=1}^m W_i \leqslant \sum_{i=m+1}^{2m} W_i\right) \leqslant \exp\left[-m\{1 + \eta_1(\alpha_0, z_0)\}I(\alpha_0, z_0)\right],$$

in which $\eta_1(\alpha_0, z_0) = O(1/\{mI(\alpha_0, z_0)\}^{1/2})$. By (35) and (10) of Assumptions 1, we have $1/\{mI(\alpha_0, z_0)\}^{1/2} \leqslant 1/(me^{2\overline{\alpha}}\underline{C})^{1/2} \to 0$. Then $-\eta_1(\alpha_0, z_0) \leqslant \epsilon/2$ for all sufficiently large n. Therefore,

$$\mathbb{P}\left(\sum_{i=1}^m W_i \leqslant \sum_{i=m+1}^{2m} W_i\right) \leqslant \exp\left\{-m\left(1 - \frac{\epsilon}{2}\right)I(\alpha_0, z_0)\right\}.$$

Since $z_0 \sim N(\mu, \tau^2 I)$ under H_0, we have $\|z_0 - \mu\|_2^2/\tau^2 \sim \chi^2(d)$. Since $\tau \to 0$ as $n \to \infty$, the inequality below holds for all sufficiently large n:

$$\left(1 - \frac{\epsilon}{4}\right)\frac{\rho^2}{\tau^2} \geqslant d + 2\left\{d\left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\}^{1/2} + \left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{\tau^2}.$$

Then by Lemma 1 of Laurent and Massart (2000), we can get

$$\begin{aligned}
\mathbb{P}_{H_0}(z_0 \in \mathcal{B}_\epsilon^c) =& \mathbb{P}_{H_0}\left\{\frac{1}{\tau^2}\|z_0 - \mu\|_2^2 > \left(1 - \frac{\epsilon}{4}\right)\frac{\rho^2}{\tau^2}\right\} \\
\leqslant& \mathbb{P}_{H_0}\left[\frac{1}{\tau^2}\|z_0 - \mu\|_2^2 \geqslant d + 2\left\{d\left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\}^{1/2} + \left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{\tau^2}\right] \\
\leqslant& \exp\left\{-\left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\}. \qquad (37)
\end{aligned}$$

Therefore by (36),

$$\begin{aligned}
&\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-}) \\
&\leqslant \mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-m\left(1 - \frac{\epsilon}{2}\right)I(\alpha_0, z_0)\right\}\right] + \mathbb{P}_{H_0}(z_0 \in \mathcal{B}_\epsilon^c) \\
&\leqslant \mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-m\left(1 - \frac{\epsilon}{2}\right)I(\alpha_0, z_0)\right\}\right] + \exp\left\{-\left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\}. \qquad (38)
\end{aligned}$$

Combining (38) with the second inequality of (22), we get

$$\begin{aligned}
\nu_n \leqslant& 2\mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-m\left(1 - \frac{\epsilon}{2}\right)I(\alpha_0, z_0)\right\}\right] + 2\exp\left\{-\left(1 - \frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\} \\
\leqslant& \mathbb{E}_{H_0}^{\alpha_0, z_0}\left[1(z_0 \in \mathcal{B}_\epsilon)\exp\left\{-m(1 - \epsilon)I(\alpha_0, z_0)\right\}\right] + \exp\left\{-(1 - \epsilon)\frac{\rho^2}{2\tau^2}\right\}.
\end{aligned}$$

Here the last inequality holds because $\epsilon/2 > \log 2/(me^{2\overline{\alpha}}\underline{C}) \geqslant \log 2/\{mI(\alpha_0, z_0)\}$ by (10) of Assumption 1 and $\epsilon/2 > (2\tau^2 \log 2)/\rho^2$ for all sufficiently large n.

Lower bound For the lower bound, when $z_0 \in \mathcal{B}_\epsilon$, we apply the Chernoff argument in Gao et al. (2017) and Gao and Ma (2020) to get

$$\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0) \geqslant \exp\left[-m\{1 + \eta_2(\alpha_0, z_0)\}I(\alpha_0, z_0)\right].$$

in which $\eta_2(\alpha_0, z_0) = O(1/\{mI(\alpha_0, z_0)\}^{1/2})$. By (35) and (10) of Assumption 1, we get $\eta_2(\alpha_0, z_0) \leqslant \epsilon$ for all sufficiently large n. Therefore,

$$\mathbb{P}\left(\sum_{i=1}^{m} W_i \leqslant \sum_{i=m+1}^{2m} W_i\right) \geqslant \exp\left\{-m(1 + \epsilon)I(\alpha_0, z_0)\right\}.$$

It is clear that $\mathcal{L}_- \subset \mathcal{B}_\epsilon^c$. When $z_0 \in \mathcal{L}_-$, we have $p(\alpha_0, z_0) < q(\alpha_0, z_0)$, so

$$\mathbb{P}(A_{0,+} \leqslant A_{0,-} \mid \alpha_0, z_0) \geqslant \frac{1}{2}.$$

Also,

$$\begin{aligned}\mathbb{P}_{H_0}(z_0 \in \mathcal{L}_-) =& \mathbb{P}_{H_0}\left\{(z_0-\mu)^\top H\mu < -\mu^\top H\mu\right\}\\ =& \Phi\left(-\frac{\mu^\top H\mu}{\tau(\mu^\top H^2\mu)^{1/2}}\right) \geqslant \exp\left\{-\left(1+\frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\},\end{aligned}$$

where the last inequality is due to Mill's ratio. Therefore, by (36) again,

$$\begin{aligned}&\mathbb{P}_{H_0}(A_{0,+} \leqslant A_{0,-})\\ &\geqslant \mathbb{E}_{H_0}^{\alpha_0,z_0}\left[1(z_0\in\mathcal{B}_\epsilon)\exp\left\{-m(1+\epsilon)I(\alpha_0,z_0)\right\}\right] + \frac{1}{2}\mathbb{P}_{H_0}(z_0\in\mathcal{L}_-)\\ &\geqslant \mathbb{E}_{H_0}^{\alpha_0,z_0}\left[1(z_0\in\mathcal{B}_\epsilon)\exp(-m(1+\epsilon)I(\alpha_0,z_0))\right] + \frac{1}{2}\exp\left\{-\left(1+\frac{\epsilon}{2}\right)\frac{\rho^2}{2\tau^2}\right\}\\ &\geqslant \mathbb{E}_{H_0}^{\alpha_0,z_0}\left[1(z_0\in\mathcal{B}_\epsilon)\exp\left\{-m(1+\epsilon)I(\alpha_0,z_0)\right\}\right] + \exp\left\{-(1+\epsilon)\frac{\rho^2}{2\tau^2}\right\}. \qquad (39)\end{aligned}$$

Here the last inequality holds because $\epsilon/2 \geqslant (2\tau^2\log 2)/\rho^2$ for sufficiently large n. Combining (39) and the first inequality in (22), we obtain the first inequality in (20).

Appendix B. Proof of Proposition 7

The following lemma will be useful in the proof of Proposition 7.

Lemma 11 *Suppose a d-dimensional random vector $z \sim N(\mu, \tau^2 I_d)$. Let M be a positive constant. Conditional on the event $\|z-\mu\|_2 \leqslant \eta$ with $\eta/\tau \to \infty$ and $\tau \to 0$, we have, for $\|t\|_2 \leqslant M$,*

$$\widetilde{\mathbb{E}}\{\exp(z^\top t)\} = \exp\left(\mu^\top t + \frac{\tau^2 t^\top t}{2}\right)\{1-o(1)\},$$

where C is a constant and $\widetilde{\mathbb{E}}$ denotes the expectation taken over the conditional measure of z on $\|z-\mu\|_2 \leqslant \eta$.

Proof Without loss of generality, we assume $\mu = 0$. We calculate

$$\begin{aligned}\widetilde{\mathbb{E}}\{\exp(z^\top t)\} &= \frac{\int_{\|z\|_2\leqslant\eta}\exp(z^\top t)\exp\{-z^\top z/(2\tau^2)\}/(2\pi\tau^2)^{d/2}dz}{\int_{\|z\|_2\leqslant\eta}\exp\{-z^\top z/(2\tau^2)\}/(2\pi\tau^2)^{d/2}dz}\\ &= \exp(\tau^2 t^\top t/2)\frac{\int_{\|z+\tau t\|_2\leqslant\eta/\tau}\exp(-z^\top z/2)/(2\pi)^{d/2}dz}{\int_{\|z\|_2\leqslant\eta/\tau}\exp(-z^\top z/2)/(2\pi)^{d/2}dz}.\end{aligned}$$

Denote the probability measure of $N(0, I_d)$ by $\mathbb{P}_0$ and we define

$$\begin{aligned}A &= \int_{\|z+\tau t\|_2\leqslant\eta/\tau}\exp(-z^\top z/2)/(2\pi)^{d/2}dz = \mathbb{P}_0(\|z+\tau t\|_2\leqslant\eta/\tau),\\ B &= \int_{\|z\|_2\leqslant\eta/\tau}\exp(-z^\top z/2)/(2\pi)^{d/2}dz = \mathbb{P}_0(\|z\|_2\leqslant\eta/\tau) = \mathbb{P}\{\chi_d^2\leqslant(\eta/\tau)^2\}.\end{aligned}$$

We have that

$$\mathbb{P}\{\chi_d^2 \leqslant (\eta/\tau - \tau\|t\|_2)^2\} = \mathbb{P}_0(\|z\|_2 \leqslant \eta/\tau - \tau\|t\|_2) \leqslant A \leqslant \mathbb{P}_0(\|z\|_2 \leqslant \eta/\tau) = \mathbb{P}\{\chi_d^2 \leqslant (\eta/\tau)^2\}.$$

As a result, we bound

$$1 \geqslant \frac{A}{B} \geqslant \frac{\mathbb{P}\{\chi_d^2 \leqslant (\eta/\tau - \tau\|t\|_2)^2\}}{\mathbb{P}\{\chi_d^2 \leqslant (\eta/\tau)^2\}} = 1 - o(1).$$

The last equality comes from the trivial bound of χ^2 distribution after choosing η/τ sufficiently large such that

$$\frac{\mathbb{P}\{(\eta/\tau)^2 \leqslant \chi_d^2 \leqslant (\eta/\tau - \tau\|t\|)^2\}}{\mathbb{P}\{\chi_d^2 \leqslant (\eta/\tau)^2\}} \leqslant 2f_d(\eta/\tau - \tau M)\tau M \leqslant C\tau M,$$

where f_d is the density function of χ_d and $C = 2\sup_x f_d(x)$. ∎

Proof [Proof of Proposition 7] First, by law of total expectation,

$$\mathbb{P}\{\ell(\sigma, \hat{\sigma}^0) > \gamma\} = \mathbb{E}^{\{\alpha_i, z_i\}_{i=1}^n}\left[\mathbb{P}\left\{\ell(\sigma, \hat{\sigma}^0) > \gamma \mid \{\alpha_i, z_i\}_{i=1}^n\right\}\right].$$

Given $\{\alpha_i, z_i\}_{i=1}^n$, the probability matrix P is deterministic. Let μ_i be the mean value of z_i, that is, $\mu_i = \mu$ for $i = 1, \ldots, n_1$ and $\mu_i = -\mu$ for $i = n_1 + 1, \ldots, n$. Let $\xi_{ij} = \mathbb{E}\{\exp(z_i^\top H z_j)\}$ for $i \neq j$ and $\xi_{ii} = \mathbb{E}\{\exp(z_1^\top H z_2)\}$. Define

$$B_{ij} = e^{\alpha_i + \alpha_j}\xi_{ij}. \tag{40}$$

We further denote $\xi_+ = \mathbb{E}\{\exp(z_1^\top H z_2)\}$ and $\xi_- = \mathbb{E}\{\exp(z_1^\top H z_{n_1+1})\} = \mathbb{E}\{\exp(-z_1^\top H z_2)\}$, then $B_{ij} = e^{\alpha_i+\alpha_j}\xi_+$ if $\sigma_i = \sigma_j$ and $B_{ij} = e^{\alpha_i+\alpha_j}\xi_-$ otherwise. It is clear that B is a matrix of rank 2, and we will show the proximity of B and $\widehat{P}$ on a high-probability event.

Step 1: Finding a high probability event. Define $\mathbb{D} = \{(\omega_1, \ldots, \omega_n) : (1/n_u)\sum_{\{i:\sigma_i = u\}} e^{4\omega_i} \leqslant \overline{L}^4 \text{ for } u = 1, 2\}$. By (12) of Assumption 1,

$$\mathbb{P}\{(\omega_1, \ldots, \omega_n) \in \mathbb{D}^c\} \leqslant 2n^{-(1+C_1)} \leqslant n^{-(1+C_1/2)}. \tag{41}$$

Let $\eta = \tau(12\log n)^{1/2}$, then by Assumption 2, $\eta \leqslant 12^{1/2}c$. Define

$$\mathbb{B}_\eta = \{(z_1, \ldots, z_n) : \|z_i - \mu_i\|_2 \leqslant \eta, 1 \leqslant i \leqslant n\}.$$

Since $\eta^2/\tau^2 > d + 2\{d\eta^2/(4\tau^2)\}^{1/2} + \eta^2/(2\tau^2)$ when n is large, by Lemma 1 of Laurent and Massart (2000),

$$\begin{aligned}\mathbb{P}(\|z_i - \mu_i\|_2 > \eta) &= \mathbb{P}\left(\frac{1}{\tau^2}\|z_i - \mu_i\|_2^2 > \frac{\eta^2}{\tau^2}\right) \\ &< \mathbb{P}\left\{\frac{1}{\tau^2}\|z_i - \mu_i\|_2^2 - d > 2\left(d\frac{\eta^2}{4\tau^2}\right)^{1/2} + \frac{\eta^2}{2\tau^2}\right\} \leqslant \exp\left(-\frac{\eta^2}{4\tau^2}\right).\end{aligned}$$

Therefore,

$$\mathbb{P}(\mathbb{B}_\eta^c) \leqslant n \exp\left(-\frac{\eta^2}{4\tau^2}\right) = n^{-2}. \tag{42}$$

Assume $(z_1, \ldots, z_n) \in \mathbb{B}_\eta$, then $z_i^\top H z_j \leqslant \mu_i^\top H \mu_j + \eta \|H\mu_i\|_2 + \eta \|H\mu_j\|_2 + \eta^2 \|H\|_2 \leqslant \mu_i^\top H \mu_j + 12^{1/2} c \|H\mu_i\|_2 + 12^{1/2} c \|H\mu_j\|_2 + 12 c^2 \|H\|_2$ which is a constant. Hence there is a positive constant $\overline{\xi}$ such that on $\mathbb{B}_\eta$

$$e^{z_i^\top H z_j} \leqslant \overline{\xi}. \tag{43}$$

Let $f_{ij} = \left\{\exp(z_i^\top H z_j) - \xi_{ij}\right\}^2$, and define the set

$$\mathbb{C}_r = \left\{(z_1, \ldots, z_n) : \sum_{1 \leqslant i \neq j \leqslant n} f_{ij} \leqslant 4r^2 n(n-1)/(\log n)^{1-\epsilon_1}\right\}$$

for any small constant $\epsilon_1 \in (0, 0.01)$ and some fixed constant $r > 0$. We will specify the choice of r later. Since ξ_{ij}, η and $\|H\|_2$ are all constants, by (43), f_{ij} has a uniform constant upper bound for all $1 \leqslant i \neq j \leqslant n$ on $\mathbb{B}_\eta$, which we denote by $\overline{f}$. Write Φ_η^+ as the measure of z_i conditioned on $\|z_i - \mu\|_2 \leqslant \eta$ for $i \in [n_1]$, and Φ_η^- for $n_1 + 1 \leqslant i \leqslant n$. The conditional distribution of $\{z_i\}_{1 \leqslant i \leqslant n}$ on $\mathbb{B}_\eta$ is

$$\underbrace{\Phi_\eta^+ \times \cdots \times \Phi_\eta^+}_{n_1} \times \underbrace{\Phi_\eta^- \times \cdots \times \Phi_\eta^-}_{n_2},$$

where $\times$ denotes the product measure. In particular, z_i's are still mutually independent conditioned on $\mathbb{B}_\eta$. Hence, for any particular $i \in [n]$, f_{ij} $(1 \leqslant j \leqslant n, j \neq i)$ are independent, and follow one of two distributions, depending on whether node j is in the same community as node i. Thus, we define

$$\begin{aligned}
f_{i+} &= \widetilde{\mathbb{E}}^{z_j}(f_{ij} \mid z_i) \quad (1 \leqslant j \leqslant n_1, j \neq i), \\
f_{i-} &= \widetilde{\mathbb{E}}^{z_j}(f_{ij} \mid z_i) \quad (n_1 + 1 \leqslant j \leqslant n, j \neq i), \\
f_{++} &= \widetilde{\mathbb{E}}^{z_i}(f_{i+}) \quad (1 \leqslant i \leqslant n_1), \\
f_{-+} &= \widetilde{\mathbb{E}}^{z_i}(f_{i+}) \quad (n_1 + 1 \leqslant i \leqslant n), \\
f_{+-} &= \widetilde{\mathbb{E}}^{z_i}(f_{i-}) \quad (1 \leqslant i \leqslant n_1), \\
f_{--} &= \widetilde{\mathbb{E}}^{z_i}(f_{i-}) \quad (n_1 + 1 \leqslant i \leqslant n),
\end{aligned}$$

where $\widetilde{\mathbb{E}}^{z_j}(\cdot \mid z_i)$ in the first two equations denotes expectation with respect to the distribution of z_j conditional on z_i and $\|z_j - \mu_j\|_2 \leqslant \eta$, and $\widetilde{\mathbb{E}}^{z_i}(\cdot)$ in the last four equalities means expectation with respect to the distribution of z_i conditional on $\|z_i - \mu_i\|_2 \leqslant \eta$. By Bernstein's inequality, we obtain

$$\begin{aligned}
&\widetilde{\mathbb{P}}\left\{\sum_{j \neq i}^n f_{ij} - \sum_{j \neq i}^n \widetilde{\mathbb{E}}^{z_j}(f_{ij} \mid z_i) > r^2 \frac{n-1}{(\log n)^{1-\epsilon_1}} \mid z_i\right\} \\
&\leqslant \exp\left\{-\frac{r^4 (n-1)^2/(\log n)^{2(1-\epsilon_1)}}{2\sum_{j \neq i} \widetilde{\operatorname{Var}}^{z_j}(f_{ij} \mid z_i) + \frac{2}{3}\overline{f} r^2 (n-1)/(\log n)^{1-\epsilon_1}}\right\},
\end{aligned}$$

where $\widetilde{\mathbb{P}}$ and $\widetilde{\mathrm{Var}}^{z_j}(\cdot)$ are taken over the distribution of z_j conditional on $\|z_j - \mu_j\|_2 \leqslant \eta$. By direct calculation we have

$$\widetilde{\mathrm{Var}}^{z_j}(f_{ij} \mid z_i) = \widetilde{M}_{ij}^{(4)} - 4\xi_{ij}\widetilde{M}_{ij}^{(3)} + 4\xi_{ij}^2\widetilde{M}_{ij}^{(2)} + 4\xi_{ij}\widetilde{M}_{ij}^{(1)}\widetilde{M}_{ij}^{(2)} - (\widetilde{M}_{ij}^{(2)})^2 - 4\xi_{ij}^2(\widetilde{M}_{ij}^{(1)})^2,$$

where $\widetilde{M}_{ij}^{(l)} = \widetilde{\mathbb{E}}^{z_j}\{\exp(l z_i^\top H z_j) \mid z_i\}$. Let $\zeta_{ij} = z_i^\top H\mu_j$ and $\iota_i = z_i^\top H^2 z_i$. Since $\|Hz_i\|_2$ is upper bounded by a constant, by Lemma 11, $\widetilde{M}_{ij}^{(l)} = \exp\left(l\zeta_{ij} + \tau^2 l^2 \iota_i/2\right)\{1 - o(1)\}$. Further calculation leads to

$$\begin{aligned}&\widetilde{\mathrm{Var}}^{z_j}(f_{ij} \mid z_i)\\ =&(e^{\tau^2\iota_i} - 1)e^{2\zeta_{ij}+\tau^2\iota_i}\left\{e^{2\zeta_{ij}+3\tau^2\iota_i}(e^{\tau^2\iota_i}+1)(e^{2\tau^2\iota_i}+1) - 4\xi_{ij}e^{\zeta_{ij}+\frac{3}{2}\tau^2\iota_i}(e^{\tau^2\iota_i}+1) + 4\xi_{ij}^2\right\}\{1 + o(1)\},\end{aligned}$$

which is upper bounded by $c_1\tau^2$ with some constant $c_1 > 0$, since ξ_{ij}, ζ_{ij} and ι_i are upper bounded by constants. By Assumption 2, we have $2\sum_{j\neq i}\widetilde{\mathrm{Var}}^{z_j}(f_{ij} \mid z_i) \leqslant 2c^2c_1(n-1)/\log n \leqslant \overline{f}r^2(n-1)/\{3(\log n)^{1-\epsilon_1}\}$ for large n. Consequently,

$$\widetilde{\mathbb{P}}\left\{\sum_{j\neq i}^n f_{ij} - \sum_{j\neq i}^n \widetilde{\mathbb{E}}^{z_j}(f_{ij} \mid z_i) > r^2\frac{n-1}{(\log n)^{1-\epsilon_1}} \mid z_i\right\} \leqslant \exp\left\{-\frac{r^2(n-1)}{\overline{f}(\log n)^{1-\epsilon_1}}\right\} \leqslant n^{-(2+C_2)} \quad (44)$$

for some constant $C_2 > 0$.

Recall that

$$\sum_{j\neq i}^n \widetilde{\mathbb{E}}^{z_j}(f_{ij} \mid z_i) = \begin{cases}(n_1-1)f_{i+} + n_2 f_{i-}, & 1 \leqslant i \leqslant n_1,\\ n_1 f_{i+} + (n_2-1)f_{i-}, & n_1+1 \leqslant i \leqslant n.\end{cases}$$

Since $f_{i+}, f_{i-} \leqslant \overline{f}$ on $\mathbb{B}_\eta$ for any $1 \leqslant i \leqslant n$, by Bernstein's inequality again, we obtain

$$\widetilde{\mathbb{P}}\left\{\sum_{i=1}^{n_1} f_{i+} - n_1 f_{++} > r^2\frac{n_1}{(\log n)^{1-\epsilon_1}}\right\} \leqslant \exp\left\{-\frac{r^4 n_1^2/(\log n)^{2(1-\epsilon_1)}}{2n_1\widetilde{\mathrm{Var}}^{z_1}(f_{1+}) + \frac{2}{3}\overline{f}r^2 n_1/(\log n)^{1-\epsilon_1}}\right\}.$$

We further bound the right hand side of the above display. By definition, we have

$$f_{1+} = \widetilde{\mathbb{E}}^{z_j}(f_{1j} \mid z_1) = \widetilde{M}_{1j}^{(2)} - 2\xi_+\widetilde{M}_{1j}^{(1)} + \xi_+^2 \quad (1 \leqslant j \leqslant n_1),$$

the variance of which is $\widetilde{\mathrm{Var}}^{z_1}(\widetilde{M}_{1j}^{(2)}) + 4\xi_+^2\widetilde{\mathrm{Var}}^{z_1}(\widetilde{M}_{1j}^{(1)}) - 4\xi_+\widetilde{\mathrm{Cov}}^{z_1}(\widetilde{M}_{1j}^{(2)}, \widetilde{M}_{1j}^{(1)})$. Since z_1 is bounded by constants and $\tau^2 \to 0$, we can find a constant $c_1' > 0$ such that $1 \leqslant \exp(4\tau^2\iota_1) \leqslant 1 + c_1'\tau^2$. Then we get

$$\begin{aligned}\widetilde{\mathrm{Var}}^{z_1}(\widetilde{M}_{1j}^{(2)}) &= \left\{\widetilde{\mathbb{E}}^{z_1}e^{4\zeta_{1j}+4\tau^2\iota_1} - \left(\widetilde{\mathbb{E}}^{z_1}e^{2\zeta_{1j}+2\tau^2\iota_1}\right)^2\right\}\{1 + o(1)\}\\ &\leqslant 2\left\{(1 + c_1'\tau^2)\widetilde{\mathbb{E}}^{z_1}e^{4\zeta_{1j}} - \left(\widetilde{\mathbb{E}}^{z_1}e^{2\zeta_{1j}}\right)^2\right\}\\ &= 2\left\{(1 + c_1'\tau^2)e^{4\mu_1^\top H\mu_j + 8\tau^2\mu_j^\top H^2\mu_j} - e^{4\mu_1^\top H\mu_j + 4\tau^2\mu_j^\top H^2\mu_j}\right\}\{1 + o(1)\}\\ &\leqslant 4e^{4\mu_1^\top H\mu_j + 4\tau^2\mu_j^\top H^2\mu_j}\left(e^{4\tau^2\mu_j^\top H^2\mu_j} - 1 + \tau^2 c_1' e^{4\tau^2\mu_j^\top H^2\mu_j}\right)\\ &\leqslant c_2'\tau^2,\end{aligned}$$

for some constant $c_2' > 0$. The last inequality is again due to $\tau^2 \to 0$. We can use similar argument to get $\widetilde{\mathrm{Var}}^{z_1}(\widetilde{M}_{1j}^{(1)}) \leqslant c_3'\tau^2$ and $\widetilde{\mathrm{Cov}}^{z_1}(\widetilde{M}_{1j}^{(2)}, \widetilde{M}_{1j}^{(1)}) \leqslant c_4'\tau^2$. Therefore, we have $\widetilde{\mathrm{Var}}(f_{1+}) \leqslant c_2\tau^2 \leqslant c^2c_2/\log n \leqslant \overline{f}r^2/\{6(\log n)^{1-\epsilon_1}\}$, where $c_2 > 0$ is a constant. This implies

$$\widetilde{\mathbb{P}}\left\{\sum_{i=1}^{n_1} f_{i+} - n_1 f_{++} > r^2\frac{n_1}{(\log n)^{1-\epsilon_1}}\right\} \leqslant \exp\left\{-\frac{r^2 n_1}{\overline{f}(\log n)^{1-\epsilon_1}}\right\} \leqslant n^{-(3+C_3)} \tag{45}$$

for some constant $C_3 > 0$. Similarly, we also obtain

$$\widetilde{\mathbb{P}}\left\{\sum_{i=1}^{n_1} f_{i-} - n_1 f_{+-} > r^2\frac{n_1}{(\log n)^{1-\epsilon_1}}\right\} \leqslant n^{-(3+C_3)} \tag{46}$$

$$\widetilde{\mathbb{P}}\left\{\sum_{i=n_1+1}^{n} f_{i+} - n_2 f_{-+} > r^2\frac{n_2}{(\log n)^{1-\epsilon_1}}\right\} \leqslant n^{-(3+C_3)} \tag{47}$$

$$\widetilde{\mathbb{P}}\left\{\sum_{i=n_1+1}^{n} f_{i-} - n_2 f_{--} > r^2\frac{n_2}{(\log n)^{1-\epsilon_1}}\right\} \leqslant n^{-(3+C_3)}. \tag{48}$$

Next we bound f_{+-}, f_{+-}, f_{-+}, and f_{--}. Since z_1 is bounded by constants and $\tau^2 \to 0$, we can find constants $c_1'' > 0$ such that $\exp(2\tau^2\iota_1) \leqslant 1 + c_1''\tau^2$. Then

$$\begin{aligned}
f_{++} &= \widetilde{\mathbb{E}}^{z_1}\left(\widetilde{M}_{12}^{(2)} - 2\xi_+\widetilde{M}_{1j}^{(1)} + \xi_+^2\right)\\
&= \left\{\widetilde{\mathbb{E}}^{z_1}(e^{2\zeta_{12}+2\tau^2\iota_1}) - 2\xi_+\widetilde{\mathbb{E}}^{z_1}(e^{\zeta_{12}+\frac{\tau^2}{2}\iota_1}) + \xi_+^2\right\}(1+o(1))\\
&\leqslant 2\left\{(1+c_1''\tau^2)\widetilde{\mathbb{E}}^{z_1}(e^{2\zeta_{12}}) - 2\xi_+\widetilde{\mathbb{E}}^{z_1}(e^{\zeta_{12}}) + \xi_+^2\right\}\\
&= 2\left\{(1+c_1''\tau^2)e^{2\mu^\top H\mu+2\tau^2\mu^\top H^2\mu} - 2\xi_+ e^{\mu^\top H\mu+\frac{\tau^2}{2}\mu^\top H^2\mu} + \xi_+^2\right\}\{1+o(1)\} &(49)\\
&\leqslant 4\left\{(1+c_1''\tau^2)e^{2\mu^\top H\mu+2\tau^2\mu^\top H^2\mu} - 2\xi_+ e^{\mu^\top H\mu+\frac{\tau^2}{2}\mu^\top H^2\mu} + \xi_+^2\right\}. &(50)
\end{aligned}$$

Here equality (49) is due to Lemma 11. By the definition of ξ_+, we have

$$\xi_+ = \mathbb{E}^{z_1}\left\{\mathbb{E}^{z_2}\left(e^{z_1^\top H z_2} \mid z_1\right)\right\} = \mathbb{E}^{z_1}\left(e^{z_1^\top H\mu+\frac{\tau^2}{2}z_1^\top H^2 z_1}\right).$$

Let $z_1 = \mu + \tau y_1$. Direct calculation leads to

$$\begin{aligned}
\xi_+ &= e^{\mu^\top H\mu+\frac{\tau^2}{2}\mu^\top H^2\mu}\,\mathbb{E}\left\{e^{\frac{\tau^4}{2}y_1^\top H^2 y_1+\tau\mu^\top H(I+\tau^2 H)y_1}\right\}\\
&= \left\{\det(I-\tau^4H^2)\right\}^{-1/2}\exp\left[\mu^\top H\mu + \frac{\tau^2}{2}\left\{\mu^\top H^2\mu + \mu^\top H(I+\tau^2H)(I-\tau^4H^2)^{-1}(I+\tau^2H)H\mu\right\}\right].
\end{aligned}$$

By Taylor expansion, we have $\det(I-\tau^4H^2) = 1 - \tau^4\mathrm{Tr}(H^2) + o(\tau^4)$. Further, since H^2 is p.s.d., then $1 \leqslant \left\{\det(I-\tau^4H^2)\right\}^{-1/2} \leqslant 1 + c_2''\tau^2$ for some constant $c_2'' > 0$. In addition, as $\tau \to 0$,

$$\mu^\top H^2\mu + \mu^\top H(I+\tau^2H)(I-\tau^4H^2)^{-1}(I+\tau^2H)H\mu \to 2\mu^\top H^2\mu.$$

Therefore, we have

$$1 \leqslant \exp\left[\frac{\tau^2}{2}\left\{\mu^\top H^2\mu + \mu^\top H(I+\tau^2 H)(I-\tau^4 H^2)^{-1}(I+\tau^2 H)H\mu\right\}\right] \leqslant (1+c_3''\tau^2)$$

for some constant $c_3'' > 0$. Therefore, we can find a constant $c_4'' > 0$ such that

$$e^{\mu^\top H\mu} \leqslant \xi_+ \leqslant (1+c_4''\tau^2)e^{\mu^\top H\mu}.$$

Plugging this into (50), we get

$$\begin{aligned} f_{++} &\leqslant 4\left\{(1+c_1''\tau^2)e^{2\mu^\top H\mu+2\tau^2\mu^\top H^2\mu} - 2e^{\mu^\top H\mu}\, e^{\mu^\top H\mu+\frac{\tau^2}{2}\mu^\top H^2\mu} + (1+c_4''\tau^2)^2 e^{2\mu^\top H\mu}\right\} \\ &\leqslant 4e^{2\mu^\top H\mu}\left(e^{2\tau^2\mu^\top H^2\mu} - 2e^{\frac{\tau^2}{2}\mu^\top H^2\mu} + 1 + c_5''\tau^2\right) \\ &\leqslant c_3\tau^2, \end{aligned}$$

where $c_5'' > 0$, $c_3 > 0$ are constants. The last two inequalities are both due to $\tau^2 \to 0$. We bound f_{+-}, f_{-+}, f_{--} in similar ways. Assumption 2 then ensures that for sufficiently large values of n,

$$\max\{f_{++}, f_{+-}, f_{-+}, f_{--}\} \leqslant 2r^2/(\log n)^{1-\epsilon_1}. \tag{51}$$

In view of the decomposition

$$\begin{aligned} \sum_{1\leqslant i\neq j\leqslant n} f_{ij} =& \sum_{i=1}^{n_1}\left\{\sum_{j\neq i} f_{ij} - (n_1-1)f_{i+} - n_2 f_{i-}\right\} + \sum_{i=n_1+1}^{n}\left\{\sum_{j\neq i} f_{ij} - n_1 f_{i+} - (n_2-1)f_{i-}\right\} \\ &+ (n_1-1)\sum_{i=1}^{n_1}(f_{i+}-f_{++}) + n_2\sum_{i=1}^{n_1}(f_{i-}-f_{+-}) \\ &+ n_1\sum_{i=n_1+1}^{n}(f_{i+}-f_{-+}) + (n_2-1)\sum_{i=n_1+1}^{n}(f_{i-}-f_{--}) \\ &+ n_1(n_1-1)f_{++} + n_1 n_2 f_{+-} + n_1 n_2 f_{-+} + n_2(n_2-1)f_{--} \end{aligned}$$

and that (51) implies

$$n_1(n_1-1)f_{++} + n_1 n_2 f_{+-} + n_1 n_2 f_{-+} + n_2(n_2-1)f_{--} \leqslant \frac{2r^2 n(n-1)}{(\log n)^{1-\epsilon_1}},$$

we obtain

$$
\begin{aligned}
&\mathbb{P}\left\{(z_1,\ldots,z_n)\in\mathbb{C}_r^c \mid (z_1,\ldots,z_n)\in\mathbb{B}_\eta\right\} = \widetilde{\mathbb{P}}\left\{\sum_{1\leqslant i\neq j\leqslant n} f_{ij} > 4r^2 n(n-1)/(\log n)^{1-\epsilon_1}\right\}\\
&\leqslant \widetilde{\mathbb{P}}\left\{\sum_{1\leqslant i\neq j\leqslant n} f_{ij} - n_1(n_1-1)f_{++} - n_1n_2f_{+-} - n_1n_2f_{-+} - n_2(n_2-1)f_{--} > 2r^2n(n-1)/(\log n)^{1-\epsilon_1}\right\}\\
&\leqslant \sum_{i=1}^{n_1}\widetilde{\mathbb{E}}^{z_i}\left[\widetilde{\mathbb{P}}\left\{\sum_{j\neq i} f_{ij} - (n_1-1)f_{i+} - n_2 f_{i-} > r^2(n-1)/(\log n)^{1-\epsilon_1} \mid z_i\right\}\right]\\
&+\sum_{i=n_1+1}^{n}\widetilde{\mathbb{E}}^{z_i}\left[\widetilde{\mathbb{P}}\left\{\sum_{j\neq i} f_{ij} - n_1 f_{i+} - (n_2-1) f_{i-} > r^2(n-1)/(\log n)^{1-\epsilon_1} \mid z_i\right\}\right]\\
&+\widetilde{\mathbb{P}}\left\{\sum_{i=1}^{n_1} f_{i+} - n_1 f_{++} > r^2 n_1/(\log n)^{1-\epsilon_1}\right\} + \widetilde{\mathbb{P}}\left\{\sum_{i=1}^{n_1} f_{i-} - n_1 f_{+-} > r^2 n_1/(\log n)^{1-\epsilon_1}\right\}\\
&+\widetilde{\mathbb{P}}\left\{\sum_{i=n_1+1}^{n} f_{i+} - n_2 f_{-+} > r^2 n_2/(\log n)^{1-\epsilon_1}\right\} + \widetilde{\mathbb{P}}\left\{\sum_{i=n_1+1}^{n} f_{i-} - n_2 f_{--} > r^2 n_2/(\log n)^{1-\epsilon_1}\right\}\\
&\leqslant n^{-(1+C_2)} + 4n^{-(3+C_3)} \leqslant n^{-(1+C_2)} + n^{-(1+C_3)}.
\end{aligned}
$$

The penultimate inequality is due to (44)–(48). We then have for large n

$$
\begin{aligned}
\mathbb{P}\big\{(z_1,\ldots,z_n)\in\mathbb{B}_\eta\cap\mathbb{C}_r^c\big\} &\leqslant \mathbb{P}\big\{(z_1,\ldots,z_n)\in\mathbb{C}_r^c \mid (z_1,\ldots,z_n)\in\mathbb{B}_\eta\big\}\\
&\leqslant n^{-(1+C_2)} + n^{-(1+C_3)}.
\end{aligned} \tag{52}
$$

Step 2: Bounding initialization error. The next part of the proof is in line with the proofs of Lemma 1 and Corollary 2 in Gao et al. (2018). Let B_i denote the ith row of B, which is defined by (40), and define $\bar{B}_i = \|B_i\|_1^{-1}B_i$. Throughout this part, we conduct all the calculation on the intersection of the events $\{(z_1,\ldots,z_n)\in\mathbb{B}_\eta\cap\mathbb{C}_r^c\}$ and $\{(\omega_1,\ldots,\omega_n)\in\mathbb{D}\}$.

Step 2.1: Establishing the separation condition for the rows of $\bar{B}$. Since $\bar{B}_i = \bar{B}_j$ when $\sigma_i=\sigma_j$, we only need to lower bound $\|\bar{B}_1-\bar{B}_n\|_1$. Let $L_u=\sum_{\sigma_i=u}e^{\omega_i}$ for $u=1,2$. When $L_1\xi_+ + L_2\xi_- \leqslant L_1\xi_- + L_2\xi_+$, we have

$$
\begin{aligned}
\|\bar{B}_1-\bar{B}_n\|_1 \geqslant& \sum_{i=1}^{n_1}|\bar{B}_{1i}-\bar{B}_{ni}| = \sum_{i=1}^{n_1}\left|\frac{e^{\omega_i}\xi_+}{L_1\xi_+ + L_2\xi_-} - \frac{e^{\omega_i}\xi_-}{L_1\xi_- + L_2\xi_+}\right|\\
=&\frac{1}{L_1\xi_- + L_2\xi_+}\sum_{i=1}^{n_1}e^{\omega_i}\left|\frac{L_1\xi_- + L_2\xi_+}{L_1\xi_+ + L_2\xi_-}\xi_+ - \xi_-\right|\\
\geqslant&\frac{L_1(\xi_+-\xi_-)}{L_1\xi_+ + L_2\xi_-}.
\end{aligned}
$$

Since

$$
L_u \leqslant (n_u\sum_{i=\sigma_u}e^{2\omega_i})^{1/2} \leqslant \left\{n_u(n_u\sum_{i=\sigma_u}e^{4\omega_i})^{1/2}\right\}^{1/2} \leqslant n_u\overline{L}
$$

for $u = 1, 2$, and $L_1 \geqslant n_1 e^{-\underline{\omega}} \geqslant n e^{-\underline{\omega}}/3$, we obtain

$$\|\bar{B}_1 - \bar{B}_n\|_1 \geqslant \frac{\frac{1}{3} n e^{-\underline{\omega}}(\xi_+ - \xi_-)}{n\overline{L}\xi_+} = \frac{\xi_+ - \xi_-}{3e^{\underline{\omega}}\overline{L}\xi_+}.$$

A similar argument holds when $L_1\xi_+ + L_2\xi_- > L_1\xi_- + L_2\xi_+$ by using $\|\bar{B}_1 - \bar{B}_n\|_1 \geqslant \sum_{i=n_1+1}^{n} |\bar{B}_{1i} - \bar{B}_{ni}|$ at the beginning of the sequence of inequalities. Therefore, the separation condition holds for $\bar{B}$

$$\min_{\sigma_i \neq \sigma_j} \|\bar{B}_i - \bar{B}_j\|_1 \geqslant \frac{\xi_+ - \xi_-}{3e^{\underline{\omega}}\overline{L}\xi_+}.$$

Step 2.2: Bounding $\sum_{\hat{\sigma}_i^0 \neq \sigma_i} e^{\omega_i}$. Let $\widehat{v}_1$ and $\widehat{v}_2$ be the centroids from the k-median step of Algorithm 1. Recall $J_0 = \{i : \widehat{\sigma}_i = 0\}$ from Algorithm 1. Fill matrix $\widehat{V} \in \mathbb{R}^{n \times n}$ with $\widehat{V}_i = \widehat{v}_{\hat{\sigma}_i^0}$ being its ith row, if $i \in J_0^c$ and $\widehat{V}_i = (0, \ldots, 0)$ if $i \in J_0$. Let $J = \{i \in J_0^c : \|\widehat{V}_i - \bar{B}_i\|_1 \geqslant (\xi_+ - \xi_-)/(6e^{\underline{\omega}}\overline{L}\xi_+)\}$. As in Lemma 5 of Gao et al. (2018) we define

$$\begin{aligned}
\mathcal{C}_u &= \{i \in J_0^c : \sigma_i = u, \|\widehat{V}_i - \bar{B}_i\|_1 < (\xi_+ - \xi_-)/(6e^{\underline{\omega}}\overline{L}\xi_+)\}, \\
R_1 &= \{u \in \{1, 2\} : \mathcal{C}_u = \varnothing\}, \\
R_2 &= \{u \in \{1, 2\} : \mathcal{C}_u \neq \varnothing, \text{ for all } i, j \in \mathcal{C}_u, \hat{\sigma}_i^0 = \hat{\sigma}_j^0\}, \\
R_3 &= \{u \in \{1, 2\} : \mathcal{C}_u \neq \varnothing, \text{ there exist } i, j \in \mathcal{C}_u, \text{s.t. } i \neq j, \hat{\sigma}_i^0 \neq \hat{\sigma}_j^0\}.
\end{aligned}$$

The counting argument in Lemma 5 of Gao et al. (2018) implies $|R_3| \leqslant |R_1|$. Therefore,

$$\sum_{i \in \cup_{u \in R_3} \mathcal{C}_u} e^{\omega_i} \leqslant |R_3| n\overline{L} \leqslant |R_1| n\overline{L} \leqslant 3e^{\underline{\omega}}\overline{L} \sum_{i \in J} e^{\omega_i}.$$

Here the last inequality holds because $\sum_{i \in J} e^{\omega_i} \geqslant \sum_{u \in R_1} \sum_{i \in C_u^c} e^{\omega_i} = \sum_{u \in R_1} \sum_{\sigma_i = u} e^{\omega_i} \geqslant |R_1| n e^{-\underline{\omega}}/3$. Hence, we have obtained

$$\sum_{\hat{\sigma}_i^0 \neq \sigma_i} e^{\omega_i} \leqslant \sum_{i \in J_0} e^{\omega_i} + \sum_{i \in J} e^{\omega_i} + \sum_{i \in \cup_{u \in R_3} \mathcal{C}_u} e^{\omega_i} \leqslant \sum_{i \in J_0} e^{\omega_i} + (1 + 3e^{\underline{\omega}}\bar{L}) \sum_{i \in J} e^{\omega_i}. \tag{53}$$

Step 2.3: Bounding $\sum_{i \in J_0} e^{\omega_i}$ and $\sum_{i \in J} e^{\omega_i}$. By definition of $\widehat{P}$ from Algorithm 1, we have

$$\sum_{i=1}^{n} \|\widehat{P}_i\|_1 \|\widehat{V}_i - \tilde{P}_i\|_1 \leqslant (1 + \varepsilon) \sum_{i=1}^{n} \|\widehat{P}_i\|_1 \|\bar{B}_i - \tilde{P}_i\|_1.$$

Then a bound for $\sum_{i\in J}\|\widehat{P}_i\|_1$ can be established by

$$\begin{aligned}
\sum_{i\in J}\|\widehat{P}_i\|_1 &\leqslant \frac{6e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}\sum_{i\in J}\|\widehat{P}_i\|_1\|\widehat{V}_i-\bar{B}_i\|_1\\
&\leqslant \frac{6e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}\sum_{i\in J}\left(\|\widehat{P}_i\|_1\|\widehat{V}_i-\tilde{P}_i\|_1+\|\widehat{P}_i\|_1\|\tilde{P}_i-\bar{B}_i\|_1\right)\\
&\leqslant (2+\varepsilon)\frac{6e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}\sum_{i=1}^n\|\widehat{P}_i\|_1\|\tilde{P}_i-\bar{B}_i\|_1\\
&\leqslant (2+\varepsilon)\frac{6e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}\sum_{i=1}^n 2\|\widehat{P}_i-B_i\|_1\frac{\|\widehat{P}_i\|_1}{\|\widehat{P}_i\|_1\vee\|B_i\|_1}\\
&\leqslant (2+\varepsilon)\frac{12e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}\sum_{i=1}^n\|\widehat{P}_i-B_i\|_1\\
&\leqslant (2+\varepsilon)\frac{12e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}n\|\widehat{P}-B\|_{\mathrm{F}},
\end{aligned}$$

where $\vee$ means the larger of two quantities. Since $\|B_i\|_1=e^{\alpha_i}\sum_{j=1}^n e^{\alpha_j}\xi_{ij}\geqslant e^{\omega_i}ne^{2\overline{\alpha}-\underline{\omega}}\xi_+/3$, we can bound $\sum_{i\in J}e^{\omega_i}$ by

$$\begin{aligned}
\sum_{i\in J}e^{\omega_i} &\leqslant \frac{3}{ne^{2\overline{\alpha}-\underline{\omega}}\xi_+}\sum_{i\in J}\|B_i\|_1\leqslant\frac{3}{ne^{2\overline{\alpha}-\underline{\omega}}\xi_+}\sum_{i\in J}\left(\|\widehat{P}_i\|_1+\|\widehat{P}_i-B_i\|_1\right)\\
&\leqslant \frac{3}{ne^{2\overline{\alpha}-\underline{\omega}}\xi_+}\left\{(2+\varepsilon)\frac{12e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}n\|\widehat{P}-B\|_{\mathrm{F}}+n\|\widehat{P}-B\|_{\mathrm{F}}\right\}\\
&=\frac{3}{e^{2\overline{\alpha}-\underline{\omega}}\xi_+}\left\{(2+\varepsilon)\frac{12e^{\underline{\omega}}\overline{L}\xi_+}{\xi_+-\xi_-}+1\right\}\|\widehat{P}-B\|_{\mathrm{F}}. \qquad (54)
\end{aligned}$$

We also bound $\sum_{i\in J_0}e^{\omega_i}$ by

$$\sum_{i\in J_0}e^{\omega_i}\leqslant\frac{3}{ne^{2\overline{\alpha}-\underline{\omega}}\xi_+}\sum_{i\in J_0}\|B_i\|_1\leqslant\frac{3}{ne^{2\overline{\alpha}-\underline{\omega}}\xi_+}\sum_{i\in J_0}\|\widehat{P}_i-B_i\|_1\leqslant\frac{3}{e^{2\overline{\alpha}-\underline{\omega}}\xi_+}\|\widehat{P}-B\|_{\mathrm{F}}. \qquad (55)$$

Combining (53), (54) and (55), we obtain

$$\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}}e^{\omega_i}\leqslant\sum_{i\in J_0}e^{\omega_i}+(1+3e^{\underline{\omega}}\bar{L})\sum_{i\in J}e^{\omega_i}\leqslant C'e^{3\underline{\omega}}\overline{L}^2e^{-2\overline{\alpha}}\|\widehat{P}-B\|_{\mathrm{F}} \qquad (56)$$

for some constant $C'>0$.

Step 2.4: Bounding $\|\widehat{P}-B\|_F$. We follow the argument of Lemma 6 in Gao et al. (2018). By definition of $\widehat{P}$, $\|\widehat{P}-A\|_{\mathrm{F}}^2\leqslant\|B-A\|_{\mathrm{F}}^2$. Then

$$\begin{aligned}
\|\widehat{P}-B\|_{\mathrm{F}}^2 &= \|\widehat{P}-A\|_{\mathrm{F}}^2-\|B-A\|_{\mathrm{F}}^2-2\langle\widehat{P}-B,B-A\rangle\\
&\leqslant 2|\langle\widehat{P}-B,B-A\rangle|\leqslant 2\|\widehat{P}-B\|_{\mathrm{F}}\sup_{K:\|K\|_{\mathrm{F}}=1:\mathsf{rank}(K)\leqslant 4}|\langle K,A-B\rangle|\\
&\leqslant\frac{1}{4}\|\widehat{P}-B\|_{\mathrm{F}}^2+4\sup_{K:\|K\|_{\mathrm{F}}=1:\mathsf{rank}(K)\leqslant 4}|\langle K,A-B\rangle|^2.
\end{aligned}$$

By rearranging terms we obtain

$$\|\widehat{P}-B\|_{\mathrm{F}}^2 \leqslant \frac{16}{3} \sup_{K:\|K\|_{\mathrm{F}}=1:\mathsf{rank}(K)\leqslant 4} |\langle K, A-B\rangle|^2.$$

Suppose K has singular value decomposition $K=\sum_{l=1}^4 \lambda_l u_l u_l^\top$, then

$$|\langle K, A-B\rangle| \leqslant \sum_{l=1}^4 |\lambda_l||u_l^\top (A-B)u_l| \leqslant \|A-B\|_2 \sum_{l=1}^4 |\lambda_l| \leqslant 2\,\|A-B\|_2\,.$$

Therefore, we have

$$\|\widehat{P}-B\|_{\mathrm{F}} \leqslant \frac{8}{3^{1/2}}\,\|A-B\|_2\,. \tag{57}$$

Define $Q_{ij}=\exp(\alpha_i+\alpha_j+z_i^\top Hz_j)$ for $1\leqslant i\neq j\leqslant n$ and $Q_{ii}=0$ for $1\leqslant i\leqslant n$. By the triangle inequality,

$$\|A-B\|_2 \leqslant \|A-P\|_2 + \|P-Q\|_2 + \|Q-B\|_2\,. \tag{58}$$

We bound the three terms on the right hand side separately. First by Example 4.1 in Latała et al. (2018), for any $u\geqslant 1$ and $t>0$, we bound

$$\mathbb{P}\big\{\|A-P\|_2 > 2e^{1/(2u)}b^{1/2} + C_4 e^{1/u}(u\log n)^{1/2} + t \mid P\big\} < \exp\Big(-\frac{t^2}{C_4}\Big) \tag{59}$$

with some constant $C_4>0$, where $b=\max_i\sum_{j=1}^n P_{ij}$. Observe that $\sum_{j=1}^n P_{ij}=e^{\alpha_i}\sum_{j=1}^n e^{\alpha_j}\exp(z_i^\top Hz_j)\leqslant e^{2\overline{\alpha}+\omega'}\sum_{j=1}^n e^{\omega_j}\overline{\xi}\leqslant \overline{\xi}\overline{L}ne^{2\overline{\alpha}+\omega'}$ for all $i\in[n]$. Take $t=\{C_4(1+C_4)\log n\}^{1/2}$ in (59), then conditional on P, with probability at least $1-n^{-(1+C_4)}$,

$$\|A-P\|_2 \leqslant C_1'(\overline{L}ne^{2\overline{\alpha}+\omega'})^{1/2} + C_2'(\log n)^{1/2} \tag{60}$$

for constants $C_1'>0$ and $C_2'>0$.

By definition, for $i\neq j$,

$$|P_{ij}-Q_{ij}| = e^{2\overline{\alpha}+\omega_i+\omega_j+z_i^\top Hz_j}\frac{e^{2\overline{\alpha}+\omega_i+\omega_j+z_i^\top Hz_j}}{1+e^{2\overline{\alpha}+\omega_i+\omega_j+z_i^\top Hz_j}} \leqslant e^{4\overline{\alpha}+2\omega_i+2\omega_j+2z_i^\top Hz_j} \leqslant e^{4\overline{\alpha}}e^{2\omega_i+2\omega_j}\overline{\xi}^2,$$

and $P_{ii}-Q_{ii}=0$. Then we obtain

$$\|P-Q\|_2 \leqslant \|P-Q\|_{\mathrm{F}} \leqslant \left(\sum_{i,j=1}^n e^{8\overline{\alpha}}e^{4\omega_i+4\omega_j}\overline{\xi}^4\right)^{1/2} = e^{4\overline{\alpha}}\sum_{i=1}^n e^{4\omega_i}\overline{\xi}^2 \leqslant \overline{\xi}^2\overline{L}^4 ne^{4\overline{\alpha}}. \tag{61}$$

By definition, $(Q_{ij}-B_{ij})^2=e^{4\overline{\alpha}+2\omega_i+2\omega_j}f_{ij}$ for $i\neq j$, and $(Q_{ii}-B_{ii})^2=\exp(4\overline{\alpha}+4\omega_i+2z_i^\top Hz_i)$. By Cauchy-Schwarz inequality,

$$\sum_{1\leqslant i\neq j\leqslant n} e^{4\overline{\alpha}+2\omega_i+2\omega_j}f_{ij} \leqslant \left(\sum_{1\leqslant i\neq j\leqslant n} e^{8\overline{\alpha}+4\omega_i+4\omega_j}\right)^{1/2}\left(\sum_{1\leqslant i\neq j\leqslant n} f_{ij}^2\right)^{1/2}.$$

It is straightforward to obtain the bound

$$\left(\sum_{1\leqslant i\neq j\leqslant n} e^{8\overline{\alpha}+4\omega_i+4\omega_j}\right)^{1/2} \leqslant e^{4\overline{\alpha}}\sum_{i=1}^{n} e^{4\omega_i} \leqslant \overline{L}^4 n e^{4\overline{\alpha}}.$$

Since $f_{ij} \leqslant \overline{f}$, we have

$$\begin{aligned}\left(\sum_{1\leqslant i\neq j\leqslant n} f_{ij}^2\right)^{1/2} &\leqslant \overline{f}^{1/2}\left(\sum_{1\leqslant i\neq j\leqslant n} f_{ij}\right)^{1/2}\\ &\leqslant \overline{f}^{1/2}\left\{4r^2(n-1)n/(\log n)^{1-\epsilon_1}\right\}^{1/2}\\ &\leqslant 2r\overline{f}^{1/2}n/(\log n)^{\frac{1-\epsilon_1}{2}}.\end{aligned}$$

Hence, we obtain

$$\sum_{1\leqslant i\neq j\leqslant n}(Q_{ij}-B_{ij})^2 \leqslant 2r\overline{f}^{1/2}\overline{L}^4 n^2 e^{4\overline{\alpha}}/(\log n)^{\frac{1-\epsilon_1}{2}}.$$

On the other hand,

$$\sum_{i=1}^{n}(Q_{ii}-B_{ii})^2 = \xi_+^2\sum_{i=1}^{n} e^{4\overline{\alpha}+4\omega_i} \leqslant \xi_+^2\overline{L}^4 n e^{4\overline{\alpha}} \leqslant 2r\overline{f}^{1/2}\overline{L}^4 n^2 e^{4\overline{\alpha}}/(\log n)^{\frac{1-\epsilon_1}{2}}.$$

Then we bound $\|Q-B\|_2$ by

$$\|Q-B\|_2 \leqslant \|Q-B\|_{\mathrm{F}} \leqslant 2r^{1/2}\overline{f}^{1/4}\overline{L}^2 n e^{2\overline{\alpha}}/(\log n)^{\frac{1-\epsilon_1}{4}}. \tag{62}$$

Step 2.5: Bounding $\sum_{\hat{\sigma}_i^0\neq\sigma_i} e^{\omega_i}$. Combining (56), (57), (58), (60), (61) and (62), we obtain that conditional on P, with probability at least $1-n^{-(1+C_4)}$

$$\begin{aligned}&\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i}\\ &\leqslant \frac{8}{3^{1/2}}C' e^{3\underline{\omega}}\overline{L}^2 e^{-2\overline{\alpha}}\left\{C_1'(\overline{L}ne^{2\overline{\alpha}+\omega'})^{1/2}+C_2'(\log n)^{1/2}+\overline{\xi}^2\overline{L}^4 n e^{4\overline{\alpha}}+2r^{1/2}\overline{f}^{1/4}\overline{L}^2 n e^{2\overline{\alpha}}/(\log n)^{\frac{1-\epsilon_1}{4}}\right\}\\ &\leqslant n\left\{C_1''\frac{1}{(ne^{2\overline{\alpha}-\omega'})^{1/2}}+C_2''\frac{(\log n)^{1/2}}{ne^{2\overline{\alpha}}}+C_3''e^{2\overline{\alpha}}+C_4''r^{1/2}\frac{1}{(\log n)^{\frac{1-\epsilon_1}{4}}}\right\}\end{aligned}$$

for constants $C_1'', C_2'', C_3'', C_4'' > 0$. By (10) and (11) of Assumption 1, we have $1/(ne^{2\overline{\alpha}-\omega'})^{1/2} \to 0$ and $(\log n)^{1/2}/(ne^{2\overline{\alpha}}) \to 0$. For any $\gamma > 0$, we can then make r small enough such that $\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i} \leqslant e^{-\underline{\omega}}\gamma n$. When γ is fixed, r can still be a constant bounded away from 0.

At last, putting (41), (42) and (52) together with the conclusion from the previous paragraph, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i} > e^{-\underline{\omega}}\gamma n\right) \\
&\leqslant \mathbb{E}^{\{\alpha_i,z_i\}_{i=1}^n}\left\{\mathbb{P}\left(\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i} > e^{-\underline{\omega}}\gamma n \mid \{\alpha_i,z_i\}_{i=1}^n\right)1((z_1,\ldots,z_n)\in\mathbb{B}_\eta\cap\mathbb{C}_r,(\omega_1,\ldots,\omega_n)\in\mathbb{D})\right\} \\
&+\mathbb{P}\left\{(z_1,\ldots,z_n)\in\mathbb{B}_\eta\cap\mathbb{C}_r^c\right\}+\mathbb{P}\left\{(z_1,\ldots,z_n)\in\mathbb{B}_\eta^c\right\}+\mathbb{P}\left\{(\omega_1,\ldots,\omega_n)\in\mathbb{D}^c\right\} \\
&\leqslant n^{-(1+C_4)}+n^{-(1+C_2)}+n^{-(1+C_3)}+n^{-2}+n^{-(1+C_1/2)} \\
&< n^{-(1+2C)}
\end{aligned}$$

with $0 < C < \min\{C_1/4, C_2/2, C_3/2, C_4/2, 1/2\}$.

Since $\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i} \geqslant e^{-\underline{\omega}} n\ell(\sigma,\hat{\sigma}^0)$, we immediately get

$$\mathbb{P}\left\{\ell(\sigma,\hat{\sigma}^0) > \gamma\right\} \leqslant \mathbb{P}\left(\sum_{\{i:\hat{\sigma}_i^0\neq\sigma_i\}} e^{\omega_i} > e^{-\underline{\omega}}\gamma n\right) < n^{-(1+2C)}.$$

This completes the proof. ∎

Appendix C. Proofs of Theorems 8 and 9

C.1 Combining the initial error and edge counting

Let $\widehat{\sigma}^{(-1,0)}$ be an n-dimensional vector one obtains after line 7 of Algorithm 3. The following Proposition 12 gives an error bound for $\widehat{\sigma}^{(-1,0)}$.

Proposition 12 *Suppose that Assumptions 1, 2 and 3 hold. Let $p(\alpha_1, z_1)$ and $q(\alpha_1, z_1)$ be quantities defined in* (14) *and* (15) *respectively, and*

$$I(\alpha_1,z_1) = -2\log\left(\{p(\alpha_1,z_1)q(\alpha_1,z_1)\}^{1/2}+[\{1-p(\alpha_1,z_1)\}\{1-q(\alpha_1,z_1)\}]^{1/2}\right).$$

Assume $n_1, n_2 \in \left[(1-\delta_n)n/2, (1+\delta_n)n/2\right]$. For any $\epsilon > 0$, define $\mathcal{B}_\epsilon = \{z_1 : \|z_1-\mu\|_2 \leqslant (1-\epsilon/4)^{1/2}\rho\}$. Then there is an n_ϵ such that for all $n > n_\epsilon$,

$$\begin{aligned}
&\mathbb{P}\left(\hat{\sigma}_1^{(-1,0)}\neq\sigma_1\right) \\
&\leqslant \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left[1(z_1\in\mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1-\epsilon)I(\alpha_1,z_1)\right\}\right]+\exp\left\{-(1-\epsilon)\frac{\rho^2}{2\tau^2}\right\}+n^{-(1+C)} \qquad (63)
\end{aligned}$$

for some constant $C > 0$.

Proof We start with some notation. Let $J_u = \{i : \sigma_i = u, 2 \leqslant i \leqslant n\}$, $n_u = |J_u|$, $\widehat{J}_u = \{i : \hat{\sigma}_i^{(-1,0)} = u, 2 \leqslant i \leqslant n\}$, $m_u = |\widehat{J}_u|$ for $u \in \{1,2\}$, and $J_{u_1u_2} = \{i : \hat{\sigma}_i^{(-1,0)} = u_1, \sigma_i = u_2, 2 \leqslant i \leqslant n\}$, $m_{u_1u_2} = |J_{u_1u_2}|$ for $u_1, u_2 \in \{1,2\}$. For convenience, we suppress the superscript $(-1,0)$ from $\hat{\sigma}_i^{(-1,0)}$ in the rest of this proof.

Recall the definitions of P_{ij} in (3) and $p(\alpha, z)$ and $q(\alpha, z)$ in (14) and (15). Define events

$$\mathbb{C}_1 = \Big\{\max_{2\leqslant i\leqslant n} \|z_i - \mu_i\|_2 \leqslant \eta\Big\},$$

$$\mathbb{D}_1 = \left\{ \sum_{\{i:\hat{\sigma}_i^{(-1,0)} \neq \sigma_i\}} e^{\omega_i} \leqslant e^{-\underline{\omega}} \frac{\gamma}{2}(n-1) \right\},$$

$$\mathbb{F}_1 = \left\{ \Big| \sum_{i\in J_1} P_{1i} - n_1 p(\alpha_1, z_1)\Big| \leqslant n_1 \epsilon' p(\alpha_1, z_1) \right\} \cap \left\{ \Big| \sum_{i\in J_2} P_{1i} - n_2 q(\alpha_1, z_1)\Big| \leqslant n_2 \epsilon' q(\alpha_1, z_1) \right\},$$

$$E_1 = \mathbb{C}_1 \cap \mathbb{D}_1 \cap \mathbb{F}_1,$$

where $\eta = \tau(12\log n)^{1/2}$ as in the proof of Proposition 7, $\gamma > 0$ and $\epsilon' > 0$ are fixed constants that will be specified later. It is worth mentioning that $\mathbb{C}_1$, $\mathbb{D}_1$, $\mathbb{F}_1$ and E_1 are all measurable with respect to the σ-algebra generated by $\{\alpha_i, z_i\}_{i=1}^n$ and $A^{(-1)}$. The proof of Proposition 7 implies that $\mathbb{P}(\mathbb{C}_1) \geqslant 1-(n-1)^{-2} \geqslant 1-n^{-3/2}$ and $\mathbb{P}(\mathbb{D}_1) \geqslant 1-(n-1)^{-(1+2C_1)} \geqslant 1-n^{-(1+C_1)}$ for some constant $C_1 > 0$ that depends on γ.

Conditional on α_1 and $z_1 \in \mathcal{B}_\epsilon$, we provide a probabilistic bound for $\mathbb{F}_1$ on event $\mathbb{C}_1$. With slight abuse of notation, let $\mathbb{E}$ denote the expectation with respect to the measure of z's restricted on $\mathbb{C}_1$. When $i \in J_2$ and $\sigma_1 = 1$, we have

$$\begin{aligned}\mathbb{E}(P_{1i}^2 \mid \alpha_1, z_1) &\leqslant e^{2\alpha_1}\mathbb{E}\{\exp(2\alpha_i + 2z_1^\top H z_i) \mid \alpha_1, z_1\} \\ &= e^{2\alpha_1 + 2\overline{\alpha} - 2z_1^\top H\mu}\mathbb{E}(e^{2\omega_i})\mathbb{E}\left[\exp\{2z_1^\top H(z_i + \mu)\} \mid z_1\right] \\ &\leqslant C_1' e^{2\alpha_1 + 2\overline{\alpha}} \exp\big(\tau^2\|Hz_1\|_2^2/2\big) \\ &\leqslant C_2' e^{2\alpha_1 + 2\overline{\alpha}}(1+\tau^2),\end{aligned}$$

for n sufficiently large. The first inequality in the preceding display holds as a result of $S(x) \leqslant e^x$ for $x \in \mathbb{R}$. In the second inequality, we use Assumption 1 to bound $\mathbb{E}(e^{2w_i})$, apply Lemma 11 and consider the fact that both z_1 and z_i are bounded on $\mathbb{C}_1$ and $\{z_1 \in \mathcal{B}_\epsilon\}$. The last inequality holds for n sufficiently large as $\tau \to 0$ as $n \to \infty$. We proceed to bound P_{1i} on $\alpha_1, z_1 \in \mathcal{B}_\epsilon$

$$P_{1i} \leqslant \exp(\alpha_1 + \overline{\alpha} + \omega')\exp\big(z_1^\top H z_i\big) \leqslant C_3' \exp(\alpha_1 + \overline{\alpha} + \omega'),$$

where we again apply $S(x) \leqslant e^x$ for $x \in \mathbb{R}$ and z_1 is finite on $\{z_1 \in \mathcal{B}_\epsilon\}$. On $\mathbb{C}_1 \cap \{z_1 \in \mathcal{B}_\epsilon\}$, by Assumption 1, we bound q from below by

$$q(\alpha_1, z_1) \geqslant C_4' \exp(\alpha_1 + \overline{\alpha} - \underline{\omega}), \quad \text{for } n \text{ sufficiently large.}$$

We apply Bernstein's inequality and obtain

$$\mathbb{P}\left\{ \Big|\sum_{i\in J_2} P_{1i} - n_2 q(\alpha_1, z_1)\Big| \geqslant t \mid \alpha_1, z_1 \right\} \leqslant 2\exp\left\{ -\frac{t^2}{2n_2 C_2' e^{2\alpha_1 + 2\overline{\alpha}}(1+\tau^2) + (2/3)C_3' e^{\alpha_1 + \overline{\alpha} + \omega'} t} \right\}.$$

Take $t = n_2\epsilon' q(\alpha_1, z_1) \geqslant C_4' n_2 \epsilon' e^{\alpha_1+\overline{\alpha}-\underline{\omega}}$, and we further obtain, for some proper constants C_5' and C_2,

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{i\in J_2} P_{1i} - n_2 q(\alpha_1, z_1)\right| \geqslant n_2\epsilon' q(\alpha_1, z_1) \mid \alpha_1, z_1\right\} &\leqslant 2\exp\left\{-\frac{C_4'^2 n_2 \epsilon'^2 e^{-2\underline{\omega}}}{2C_2'(1+\tau^2) + (2/3)C_3'C_4'\epsilon' e^{\omega'-\underline{\omega}}}\right\} \\ &\leqslant 2\exp\left(-C_5' n_2 \epsilon' e^{-\underline{\omega}-\omega'}\right) \\ &\leqslant \frac{1}{2} n^{-(1+C_2)}. \end{aligned} \tag{64}$$

The second inequality in the preceding display holds as $e^{\omega'-\underline{\omega}} \gtrsim 1$ by Assumption 1. We apply (11) in Assumption 1 to obtain the last inequality. A similar argument yields that conditional on $\mathbb{C}_1$, for $z_1 \in \mathcal{B}_\epsilon$

$$\mathbb{P}\left\{\left|\sum_{i\in J_1} P_{1i} - n_1 p(\alpha_1, z_1)\right| \geqslant n_1\epsilon' p(\alpha_1, z_1) \mid \alpha_1, z_1\right\} \leqslant \frac{1}{2} n^{-(1+C_2)}. \tag{65}$$

Combining (64) and (65), we obtain that conditional on α_1 and $z_1 \in \mathcal{B}_\epsilon$,

$$\mathbb{P}(\mathbb{F}_1^c \mid \mathbb{C}_1) \leqslant n^{-(1+C_2)}.$$

Together with the probabilistic bound on $\mathbb{C}_1$, for some constant C_2'', we have conditional on α_1 and $z_1 \in \mathcal{B}_\epsilon$,

$$\begin{aligned}\mathbb{P}(\mathbb{F}_1^c) &\leqslant \mathbb{P}(\mathbb{F}_1^c \mid \mathbb{C}_1)\mathbb{P}(\mathbb{C}_1) + \mathbb{P}(\mathbb{F}_1^c \mid \mathbb{C}_1^c)\mathbb{P}(\mathbb{C}_1^c) \leqslant \mathbb{P}(\mathbb{F}_1^c \mid \mathbb{C}_1) + \mathbb{P}(\mathbb{C}_1^c) \\ &\leqslant n^{-(1+C_2)} + n^{-3/2} \leqslant n^{-(1+C_2'')}, \end{aligned} \tag{66}$$

Inspection of the above argument reveals that as long as $z_1 \in \mathcal{B}_\epsilon$, the constant C_2'' in the preceding display does not depend on α_1 and z_1, whence we obtain

$$\mathbb{P}(E_1^c \mid z_1 \in \mathcal{B}_\epsilon) \leqslant \mathbb{P}(\mathbb{C}_1^c) + \mathbb{P}(\mathbb{D}_1^c) + \mathbb{P}(\mathbb{F}_1^c) \leqslant n^{-3/2} + n^{-(1+C_1)} + n^{-(1+C_2'')} \leqslant n^{-(1+C)}, \tag{67}$$

with $0 < C < \min\{1/2, C_1, C_2''\}$. It will be useful at the end of the proof to give a probabilistic bound on E_1 without conditioning on $\{z_1 \in \mathcal{B}_\epsilon\}$

$$\mathbb{P}(E_1^c) \leqslant \mathbb{P}(E_1^c \mid z_1 \in \mathcal{B}_\epsilon) + \mathbb{P}(z_1 \in \mathcal{B}_\epsilon) \leqslant n^{-(1+C)} + \exp\left\{-(1-\epsilon/2)\frac{\rho^2}{2\tau^2}\right\}, \tag{68}$$

where the last inequality follows from (37) in Lemma 6.

Next observe that

$$\begin{aligned}&\mathbb{P}_{\{\sigma_1=1\}}(\hat{\sigma}_1 = 2 \text{ and } E_1) \\ &= \mathbb{P}_{\{\sigma_1=1\}}\left(\frac{1}{m_1}\sum_{i\in\hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in\hat{J}_2} A_{1,i} \text{ and } E_1\right) \\ &= \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left\{1(z_1\in\mathcal{B}_\epsilon)\mathbb{P}\left(\frac{1}{m_1}\sum_{i\in\hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in\hat{J}_2} A_{1,i} \text{ and } E_1 \mid \alpha_1, z_1\right)\right\} \\ &+ \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left\{1(z_1\in\mathcal{B}_\epsilon^c)\mathbb{P}\left(\frac{1}{m_1}\sum_{i\in\hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in\hat{J}_2} A_{1,i} \text{ and } E_1 \mid \alpha_1, z_1\right)\right\}. \end{aligned} \tag{69}$$

We deal with the first term in the above display. Assume $z_1 \in \mathcal{B}_\epsilon$ in the following. We then have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{m_1}\sum_{i\in\hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in\hat{J}_2} A_{1,i} \text{ and } E_1 \mid \alpha_1, z_1\right) \\
&= \mathbb{E}\left[\mathbb{E}\left\{1(E_1)\mathbf{1}\left(\frac{1}{m_1}\sum_{i\in\hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in\hat{J}_2} A_{1,i}\right) \mid \{\alpha_i, z_i\}_{i=1}^n\right\} \mid \alpha_1, z_1\right] \\
&\leqslant \mathbb{E}\left[\mathbb{E}\left\{1(E_1)\mathbf{1}\left(\frac{1}{m_1}\sum_{i\in J_{11}} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in J_{22}} A_{1,i} + \frac{1}{m_2}\sum_{i\in J_{21}} A_{1,i}\right) \mid \{\alpha_i, z_i\}_{i=1}^n\right\} \mid \alpha_1, z_1\right].
\end{aligned} \tag{70}$$

The equality holds because of the tower property of conditional expectations. We now consider the conditional expectation inside the round brackets in the preceding display. Conditional on $\{\alpha_i, z_i\}_{i=1}^n$, we define for $i \in [n]$

$$W_i \overset{\text{ind}}{\sim} \text{Bernoulli}(P_{1i}).$$

Contionally on $\{\alpha_i, z_i\}_{i=1}^n$, $(A_{1i})_{2\leqslant i\leqslant n}$ are mutually independent and independent of $A^{(-1)}$, whence we have, for any $t > 0$ measurable with respect to the σ-algebra generated by $\{\alpha_i, z_i\}_{i=1}^n$ and $A^{(-1)}$,

$$\begin{aligned}
&\mathbb{E}\left\{1(E_1)\mathbf{1}\left(\frac{1}{m_1}\sum_{i\in J_{11}} A_{1,i} \leqslant \frac{1}{m_2}\sum_{i\in J_{22}} A_{1,i} + \frac{1}{m_2}\sum_{i\in J_{21}} A_{1,i}\right) \mid \{\alpha_i, z_i\}_{i=1}^n\right\} \\
&= \mathbb{E}\left\{1(E_1)\mathbf{1}\left(\frac{1}{m_1}\sum_{i\in J_{11}} W_i \leqslant \frac{1}{m_2}\sum_{i\in J_{22}} W_i + \frac{1}{m_2}\sum_{i\in J_{21}} W_i\right) \mid \{\alpha_i, z_i\}_{i=1}^n\right\} \\
&= \mathbb{E}\left[1(E_1)\mathbb{E}\left\{\mathbf{1}\left(\frac{1}{m_1}\sum_{i\in J_{11}} W_i \leqslant \frac{1}{m_2}\sum_{i\in J_{22}} W_i + \frac{1}{m_2}\sum_{i\in J_{21}} W_i\right) \mid \{\alpha_i, z_i\}_{i=1}^n, A^{(-1)}\right\} \mid \{\alpha_i, z_i\}_{i=1}^n\right] \\
&\leqslant \mathbb{E}\left\{1(E_1)\prod_{i\in J_{22}}\left(P_{1i}e^{t/m_2} + 1 - P_{1i}\right)\prod_{i\in J_{21}}\left(P_{1i}e^{t/m_2} + 1 - P_{1i}\right)\prod_{i\in J_{11}}\left(P_{1i}e^{-t/m_1} + 1 - P_{1i}\right) \mid \{\alpha_i, z_i\}_{i=1}^n\right\} \\
&\leqslant \mathbb{E}\left[1(E_1)\exp\left\{\sum_{i\in J_{22}} P_{1i}(e^{t/m_2} - 1) + \sum_{i\in J_{21}} P_{1i}(e^{t/m_2} - 1) + \sum_{i\in J_{11}} P_{1i}(e^{-t/m_1} - 1)\right\} \mid \{\alpha_i, z_i\}_{i=1}^n\right].
\end{aligned} \tag{71}$$

The second equality in the preceding display holds by the tower property of conditional expectations and because E_1 is measurable with respect to the σ-algebra generated by $\{\alpha_i, z_i\}_{i=1}^n$ and $A^{(-1)}$. In the first inequality, we apply the Chernoff bound and consider the fact that m_1, m_2, P_{1i}'s and $(J_{u_1u_2})_{u_1,u_2\in[2]}$ are all measurable with respect to the σ-algebra generated by $\{\alpha_i, z_i\}_{i=1}^n$ and $A^{(-1)}$. The second inequality holds as $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$.

Write $p = p(\alpha_1, z_1)$ and $q = q(\alpha_1, z_1)$ as shorthands. Define the following quantities

$$K_1 = \exp\{m_2(e^{t/m_2} - 1)q + m_1(e^{-t/m_1} - 1)p\},$$
$$K_2 = \exp\left\{(e^{t/m_2} - 1)\Big(\sum_{i\in J_{22}} P_{1i} - m_2 q\Big)\right\},$$
$$K_3 = \exp\left\{(e^{-t/m_1} - 1)\Big(\sum_{i\in J_{11}} P_{1i} - m_1 p\Big)\right\},$$
$$K_4 = \exp\Big\{(e^{t/m_2} - 1)\sum_{i\in J_{21}} P_{1i}\Big\}.$$

It is clear that (71) is the same as $\mathbb{E}\big[1(E_1)K_1K_2K_3K_4 \mid \{\alpha_i, z_i\}_{i=1}^n\big]$. Set $t = m_1m_2\log(p/q)/(m_1+m_2)$. Next we deal with K_1, K_2, K_3 and K_4 separately.

Before we proceed, we mention the following useful facts. For any fixed $\gamma > 0$, we make n sufficiently large so that $\delta_n < \gamma$. Hence, $n_1, n_2 \in [(1-\gamma)n/2, (1+\gamma)n/2]$. On event $E_1 \subset \mathbb{D}_1$, we have

$$|\{i : \hat{\sigma}_i^{(-1,0)} \neq \sigma_i\}| \leqslant e^{\underline{\omega}} \sum_{\{i:\hat{\sigma}_i^{(-1,0)}\neq\sigma_i\}} e^{\omega_i} \leqslant \frac{\gamma}{2}(n-1) < \frac{\gamma}{2}n.$$

Therefore, we get $m_{12} \leqslant \gamma n/2$, $m_{21} \leqslant \gamma n/2$, and hence $m_1, m_2 \in [(1-2\gamma)n/2, (1+2\gamma)n/2]$. Furthermore, for $z_1 \in \mathcal{B}_\epsilon$, the lower bound (25) holds for $z_1^\top H\mu$. We denote $\underline{\xi} = \exp(\epsilon\mu^\top H\mu/8)$. For $z_1 \in \mathcal{B}_\epsilon$, on event E_1, both $\exp(z_1^\top Hz_i)$ and $\exp(z_1^\top H\mu)$ are bounded above by some constant $\overline{\xi}$, which is larger than 1 since $z_1^\top H\mu > 0$ when $z_1 \in \mathcal{B}_\epsilon$.

First we deal with the main term K_1. Since (26), (27), (28), (29), (30), (33) and (34) continue to hold for p and q, we obtain

$$(1-\kappa)e^{2\overline{\alpha}}e^{z_1^\top H\mu}D(\omega_1, z_1) \leqslant p \leqslant e^{2\overline{\alpha}}e^{z_1^\top H\mu}D(\omega_1, z_1), \tag{72}$$
$$(1-\kappa)e^{2\overline{\alpha}}e^{-z_1^\top H\mu}D(\omega_1, z_1) \leqslant q \leqslant e^{2\overline{\alpha}}e^{-z_1^\top H\mu}D(\omega_1, z_1), \tag{73}$$

where $\kappa = 1 - (1+\underline{\xi}^{-1})^2/4 \in (0,1)$ and $0 < e^{-2\underline{\omega}}\underline{D} \leqslant D(\omega_1, z_1) \leqslant e^{\omega'}\overline{D}$. For this particular choice of κ, we have

$$(1-\kappa)^{1/2}e^{z_1^\top H\mu} - 1 \geqslant \frac{1}{2}(1+\underline{\xi}^{-1})\underline{\xi} - 1 = \frac{1}{2}(\underline{\xi} - 1) > 0. \tag{74}$$

By direct calculation,

$$\begin{aligned} m_2q(e^{t/m_2} - 1) + m_1p(e^{-t/m_1} - 1) &= -\left\{m_1p + m_2q - (m_1+m_2)p^{\frac{m_1}{m_1+m_2}}q^{\frac{m_2}{m_1+m_2}}\right\} \\ &\leqslant -\frac{n}{2}\left\{p + q - 2\gamma(p-q) - 2\left(\frac{p}{q}\right)^\gamma (pq)^{1/2}\right\}. \end{aligned} \tag{75}$$

We aim to show that the term inside the round brackets of the last display and $-nI(\alpha_1, z_1)/2$ are close. To this end, first we have

$$\frac{p + q - 2\gamma(p-q) - 2(p/q)^\gamma (pq)^{1/2}}{(p^{1/2} - q^{1/2})^2} = 1 - 2\gamma\frac{p^{1/2} + q^{1/2}}{p^{1/2} - q^{1/2}} - 2\left\{\left(\frac{p}{q}\right)^\gamma - 1\right\}\frac{(pq)^{1/2}}{(p^{1/2} - q^{1/2})^2}$$

Using (72), (73) and (74), we obtain

$$2\gamma\frac{p^{1/2}+q^{1/2}}{p^{1/2}-q^{1/2}} \leqslant 2\gamma\frac{e^{\frac{1}{2}z_1^\top H\mu}+e^{-\frac{1}{2}z_1^\top H\mu}}{(1-\kappa)^{1/2}e^{\frac{1}{2}z_1^\top H\mu}-e^{-\frac{1}{2}z_1^\top H\mu}} = 2\gamma\frac{e^{z_1^\top H\mu}-1}{(1-\kappa)^{1/2}e^{z_1^\top H\mu}-1} \leqslant 4\gamma\frac{\overline{\xi}-1}{\underline{\xi}-1},$$

$$\left(\frac{p}{q}\right)^{\gamma}-1 \leqslant \left\{\frac{e^{z_1^\top H\mu}}{(1-\kappa)e^{-z_1^\top H\mu}}\right\}^{\gamma}-1 \leqslant \left(\frac{\overline{\xi}^2}{1-\kappa}\right)^{\gamma}-1,$$

$$\frac{(pq)^{1/2}}{(p^{1/2}-q^{1/2})^2} \leqslant \frac{1}{\{(1-\kappa)^{1/2}e^{\frac{1}{2}z_1^\top H\mu}-e^{-\frac{1}{2}z_1^\top H\mu}\}^2} = \frac{e^{z_1^\top H\mu}}{\{(1-\kappa)^{1/2}e^{z_1^\top H\mu}-1\}^2} \leqslant \frac{4\overline{\xi}}{(\underline{\xi}-1)^2}.$$

We choose γ such that the second last and third last displays are sufficiently small. Hence, for sufficiently small constant $\gamma > 0$,

$$\frac{p+q+2\gamma(p-q)-2\,(p/q)^{\gamma}\,(pq)^{1/2}}{(p^{1/2}-q^{1/2})^2} \geqslant 1-\frac{\epsilon}{4}. \tag{76}$$

We also have

$$I(\alpha_1, z_1) = -\,2\log\left[1-\frac{1}{2}(p^{1/2}-q^{1/2})^2-\frac{1}{2}\{(1-p)^{1/2}-(1-q)^{1/2}\}^2\right].$$

Let

$$\begin{aligned}\beta &= \frac{1}{2}(p^{1/2}-q^{1/2})^2+\frac{1}{2}\{(1-p)^{1/2}-(1-q)^{1/2}\}^2\\ &= \frac{1}{2}(p^{1/2}-q^{1/2})^2\left[1+\frac{(p^{1/2}+q^{1/2})^2}{\{(1-p)^{1/2}+(1-q)^{1/2}\}^2}\right].\end{aligned}$$

By (72), (73), (9) of Assumption 1, and that $\exp(z_1^\top H\mu) \leqslant \overline{\xi}$, we have $p, q \leqslant 3/4$. Thus,

$$\begin{aligned}\frac{(p^{1/2}+q^{1/2})^2}{\{(1-p)^{1/2}+(1-q)^{1/2}\}^2} &\leqslant \frac{e^{2\overline{\alpha}}\left(e^{\frac{1}{2}z_1^\top H\mu}+e^{-\frac{1}{2}z_1^\top H\mu}\right)^2 D(\omega_1, z_1)}{(1/2+1/2)^2}\\ &\leqslant e^{2\overline{\alpha}+\omega'}(\overline{\xi}^{1/2}+\underline{\xi}^{-1/2})^2\overline{D},\end{aligned}$$

which goes to 0 as $2\overline{\alpha}+\omega' \to -\infty$ by (9) of Assumption 1. Consequently,

$$\begin{aligned}\beta &\leqslant \frac{1}{2}e^{2\overline{\alpha}+\omega'}\left\{e^{\frac{1}{2}z_1^\top H\mu}-(1-\kappa)^{1/2}e^{-\frac{1}{2}z_1^\top H\mu}\right\}^2\overline{D}\left\{1+e^{2\overline{\alpha}+\omega'}\left(\overline{\xi}^{1/2}+\underline{\xi}^{-1/2}\right)^2\overline{D}\right\}\\ &\leqslant \frac{1}{2}e^{2\overline{\alpha}+\omega'}\left\{\overline{\xi}^{1/2}-\frac{1}{2}(1+\underline{\xi}^{-1})\overline{\xi}^{-1/2}\right\}^2\overline{D}\left\{1+e^{2\overline{\alpha}+\omega'}(\overline{\xi}^{1/2}+\underline{\xi}^{-1/2})^2\overline{D}\right\},\end{aligned}$$

which also goes to 0 as $2\overline{\alpha}+\omega' \to -\infty$. Since $\log(1-\beta) \geqslant -\beta-\beta^2$ for all $0<\beta<1/2$, we obtain $I(\alpha_1, z_1) \leqslant 2\beta+2\beta^2$. Therefore,

$$\frac{I(\alpha_1, z_1)}{(p^{1/2}-q^{1/2})^2} \leqslant (1+\beta)\left[1+\frac{(p^{1/2}+q^{1/2})^2}{\{(1-p)^{1/2}+(1-q)^{1/2}\}^2}\right].$$

Since the limits of β and $(p^{1/2}+q^{1/2})^2/\{(1-p)^{1/2}+(1-q)^{1/2}\}^2$ are both zeros, we have for large values of n that

$$\frac{I(\alpha_1,z_1)}{(p^{1/2}-q^{1/2})^2} \leqslant 1+\frac{\epsilon}{4}. \tag{77}$$

We combine (75), (76) and (77) to obtain

$$K_1 \leqslant \exp\left\{-\frac{n}{2}\frac{1-\epsilon/4}{1+\epsilon/4}I(\alpha_1,z_1)\right\} \leqslant \exp\left\{-\frac{n}{2}\left(1-\frac{\epsilon}{2}\right)I(\alpha_1,z_1)\right\}. \tag{78}$$

To bound K_2, we have the decomposition

$$K_2 = \exp\left[(e^{t/m_2}-1)\Big\{\sum_{i\in J_2}P_{1i}-n_2q-\sum_{i\in J_{12}}P_{1i}+(n_2-m_2)q\Big\}\right].$$

By (72) and (73), we bound $e^{t/m_2}-1$ by a constant

$$e^{t/m_2}-1=\left(\frac{p}{q}\right)^{\frac{m_1}{m_1+m_2}}-1\leqslant\left(\frac{p}{q}\right)^{\gamma+\frac{1}{2}}-1\leqslant\left\{\frac{e^{z_1^\top H\mu}}{(1-\kappa)e^{-z_1^\top H\mu}}\right\}^{\gamma+\frac{1}{2}}-1\leqslant\frac{\overline{\xi}^2}{1-\kappa}-1. \tag{79}$$

We then bound $|\sum_{i\in J_2}P_{1i}-n_2q|$, $\sum_{i\in J_{12}}P_{1i}$ and $(n_2-m_2)q$ one by one. By definition, on event E_1 we have

$$\Big|\sum_{i\in J_2}P_{1i}-n_2q\Big|\leqslant n_2\epsilon' q< n\epsilon' q. \tag{80}$$

We use $\log(1-x)\leqslant -x$ for $0<x<1$ to obtain

$$I(\alpha_1,z_1)\geqslant 2\beta=(p^{1/2}-q^{1/2})^2\left[1+\frac{(p^{1/2}+q^{1/2})^2}{\{(1-p)^{1/2}+(1-q)^{1/2}\}^2}\right].$$

By (72) and (73), we get

$$\begin{aligned}\frac{I(\alpha_1,z_1)}{(p^{1/2}-q^{1/2})^2}&\geqslant 1+\frac{1}{4}\left\{(1-\kappa)e^{2\overline{\alpha}-2\underline{\omega}}(e^{\frac{1}{2}z_1^\top H\mu}+e^{-\frac{1}{2}z_1^\top H\mu})^2\underline{D}\right\}\\&\geqslant 1+\frac{1}{4}\left\{(1-\kappa)e^{2\overline{\alpha}-2\underline{\omega}}(\underline{\xi}^{1/2}+\overline{\xi}^{-1/2})^2\underline{D}\right\}\to 1,\end{aligned} \tag{81}$$

as $\overline{\alpha}\to-\infty$. Following (74), we also have

$$\frac{q}{(p^{1/2}-q^{1/2})^2}=\frac{1}{\{(p/q)^{1/2}-1\}^2}\leqslant\frac{1}{\{(1-\kappa)^{1/2}e^{z_1^\top H\mu}-1\}^2}\leqslant\frac{4}{(\underline{\xi}-1)^2}. \tag{82}$$

Putting (79), (80), (81) and (82) together, for a suitably chosen ϵ', we obtain

$$\begin{aligned}(e^{t/m_2}-1)\Big|\sum_{i\in J_2}P_{1i}-n_2q\Big|&\leqslant\left(\frac{\overline{\xi}^2}{1-\kappa}-1\right)n\epsilon' I(\alpha_1,z_1)\frac{(p^{1/2}-q^{1/2})^2}{I(\alpha_1,z_1)}\frac{q}{(p^{1/2}-q^{1/2})^2}\\&\leqslant\frac{\epsilon}{32}nI(\alpha_1,z_1).\end{aligned} \tag{83}$$

Since $P_{1i} \leqslant \exp(\alpha_1 + \alpha_i + z_1^\top H z_i) \leqslant \bar{\xi} e^{2\bar{\alpha}+\omega_1} e^{\omega_i}$, then on event E_1 we have

$$\sum_{i \in J_{12}} P_{1i} \leqslant \bar{\xi} e^{2\bar{\alpha}+\omega_1} \sum_{i \in J_{12}} e^{\omega_i} \leqslant \bar{\xi} e^{2\bar{\alpha}+\omega_1} \sum_{\{i : \hat{\sigma}_i^{(-1,0)} \neq \sigma_i\}} e^{\omega_i} \leqslant \bar{\xi} e^{2\bar{\alpha}+\omega_1} e^{-\underline{\omega}} \frac{\gamma}{2}(n-1).$$

By (73), the definition of $D(\omega_1, z_1)$ and Assumption 1, we see $q \gtrsim e^{2\bar{\alpha}+\omega_1}$. In view of (81) and (82), we make γ small enough such that

$$\begin{aligned}
(e^{t/m_2} - 1) \sum_{i \in J_{12}} P_{1i} &\leqslant \left(\frac{\bar{\xi}^2}{1-\kappa} - 1 \right) \bar{\xi} e^{2\bar{\alpha}+\omega_1} e^{-\underline{\omega}} \frac{\gamma}{2}(n-1) \\
&\leqslant \left(\frac{\bar{\xi}^2}{1-\kappa} - 1 \right) \bar{\xi} \frac{e^{2\bar{\alpha}+\omega_1}}{q} \frac{q}{(p^{1/2} - q^{1/2})^2} \frac{(p^{1/2} - q^{1/2})^2}{I(\alpha_1, z_1)} e^{-\underline{\omega}} \frac{\gamma}{2} n I(\alpha_1, z_1) \\
&\leqslant \frac{\epsilon}{32} n I(\alpha_1, z_1).
\end{aligned} \tag{84}$$

Since $n_2 - m_2 \leqslant 3\gamma n/2$, combining (79), (81) and (82) we obtain

$$\begin{aligned}
(e^{t/m_2} - 1)(n_2 - m_2) q &\leqslant \left(\frac{\bar{\xi}^2}{1-\kappa} - 1 \right) \frac{3}{2} \gamma n I(\alpha_1, z_1) \frac{(p^{1/2} - q^{1/2})^2}{I(\alpha_1, z_1)} \frac{q}{(p^{1/2} - q^{1/2})^2} \\
&\leqslant \frac{\epsilon}{32} n I(\alpha_1, z_1)
\end{aligned} \tag{85}$$

for small enough γ. Combining (83), (84) and (85), we obtain

$$K_2 \leqslant \exp\left\{ \frac{3\epsilon}{32} n I(\alpha_1, z_1) \right\} \tag{86}$$

The same bound for K_3 is obtained similarly to bound K_2

$$K_3 \leqslant \exp\left\{ \frac{3\epsilon}{32} n I(\alpha_1, z_1) \right\}. \tag{87}$$

Lastly, the following bound for K_4 is obtained by the same argument as in establishing (84)

$$K_4 \leqslant \exp\left\{ \frac{\epsilon}{32} n I(\alpha_1, z_1) \right\}. \tag{88}$$

Combining (71), (78), (86), (87), (88), we get

$$\begin{aligned}
&\mathbb{E} \left\{ 1(E_1) \mathbf{1} \left(\frac{1}{m_1} \sum_{i \in J_{11}} A_{1,i} \leqslant \frac{1}{m_2} \sum_{i \in J_{22}} A_{1,i} + \frac{1}{m_2} \sum_{i \in J_{21}} A_{1,i} \right) \mid \{\alpha_i, z_i\}_{i=1}^n \right\} \\
&\leqslant \exp\left\{ -\frac{n}{2} (1 - \frac{15}{16}\epsilon) I(\alpha_1, z_1) \right\} \leqslant \exp\left\{ -\frac{n}{2}(1-\epsilon) I(\alpha_1, z_1) \right\}.
\end{aligned}$$

Since the rightmost side of the above display depends only on (α_1, z_1), by (70) we obtain for $z_1 \in \mathcal{B}_\epsilon$

$$\mathbb{P}\left(\frac{1}{m_1} \sum_{i \in \hat{J}_1} A_{1,i} \leqslant \frac{1}{m_2} \sum_{i \in \hat{J}_2} A_{1,i} \text{ and } E_1 \mid \alpha_1, z_1 \right) \leqslant \exp\left\{ -\frac{n}{2}(1-\epsilon) I(\alpha_1, z_1) \right\}.$$

By (69), we further have

$$\begin{aligned}\mathbb{P}_{\{\sigma_1=1\}}(\hat{\sigma}_1 = 2 \text{ and } E_1) &\leqslant \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left[1(z_1 \in \mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1-\epsilon)I(\alpha_1,z_1)\right\}\right] + \mathbb{P}_{\{\sigma_1=1\}}(z_1 \in \mathcal{B}^c_\epsilon) \\ &\leqslant \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left[1(z_1 \in \mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1-\epsilon)I(\alpha_1,z_1)\right\}\right] + \exp\left\{-(1-\epsilon/2)\frac{\rho^2}{2\tau^2}\right\},\end{aligned}$$

where the last inequality is due to (37) in Lemma 6. Finally, in view of (68), we have

$$\begin{aligned}\mathbb{P}_{\{\sigma_1=1\}}(\hat{\sigma}_1 = 2) &\leqslant \mathbb{P}_{\{\sigma_1=1\}}(\hat{\sigma}_1 = 2 \text{ and } E_1) + \mathbb{P}_{\{\sigma_1=1\}}(E_1^c) \\ &\leqslant \mathbb{E}^{\alpha_1,z_1}_{\{\sigma_1=1\}}\left[1(z_1 \in \mathcal{B}_\epsilon)\exp\left\{-\frac{n}{2}(1-\epsilon)I(\alpha_1,z_1)\right\}\right] + \exp\left\{-(1-\epsilon)\frac{\rho^2}{2\tau^2}\right\} + n^{-(1+C)}.\end{aligned}$$

■

C.2 Proof of Theorem 8

The proof strategy here is similar to that used in the proof of Theorem 2 in Gao et al. (2017). For $i \in [n]$ there is a permutation π_i such that

$$\ell(\sigma, \widehat{\sigma}^{(-i,0)}) = \frac{1}{n}\sum_{j=1}^n 1(\sigma_j \neq \pi_i(\hat{\sigma}_j^{(-i,0)})).$$

Without loss of generality, we may assume that $\pi_1 = \text{Id}$ is the identity permutation. Then by Proposition 7 and Lemma 4 in Gao et al. (2017), we obtain that for some constant $C > 0$, for each $i = 2, \ldots, n$ with probability at least $1 - n^{-(1+C)}$,

$$\widehat{\sigma}_i = \pi_i(\widehat{\sigma}_i^{(-i,0)}).$$

Together with Proposition 12, we obtain that for $i = 1, \ldots, n$,

$$\begin{aligned}\mathbb{P}(\sigma_i \neq \widehat{\sigma}_i) &\leqslant \mathbb{P}\{\sigma_i \neq \pi_i(\widehat{\sigma}_i^{(-i,0)}),\ \widehat{\sigma}_i = \pi_i(\widehat{\sigma}_i^{(-i,0)})\} + \mathbb{P}\{\widehat{\sigma}_i \neq \pi_i(\widehat{\sigma}_i^{(-i,0)})\} \\ &\leqslant \overline{\nu}_n^{\epsilon'} + 2n^{-(1+C)}. \end{aligned} \tag{89}$$

Here, for any fixed $\epsilon \in (0, 1/2)$, we pick

$$\epsilon' = \frac{\epsilon}{2}.$$

By Markov's inequality, We have

$$\begin{aligned}\mathbb{P}\left\{\ell(\sigma, \widehat{\sigma}) > \overline{\nu}_n^\epsilon\right\} &\leqslant \frac{1}{\overline{\nu}_n^\epsilon}\cdot\frac{1}{n}\sum_{i=1}^n \mathbb{P}(\sigma_i \neq \widehat{\sigma}_i) \\ &\leqslant \frac{\overline{\nu}_n^{\epsilon'}}{\overline{\nu}_n^\epsilon} + \frac{2n^{-(1+C)}}{\overline{\nu}_n^\epsilon}.\end{aligned}$$

We divide the remaining proof into two cases depending on the relative magnitude of $\overline{\nu}_n^\epsilon$ and $n^{-(1+C/2)}$.

Case 1 If $\overline{\nu}_n^\epsilon \geqslant n^{-(1+C/2)}$, then

$$\mathbb{P}\left\{\ell(\sigma,\widehat{\sigma}) > \overline{\nu}_n^\epsilon\right\} \leqslant \frac{\overline{\nu}_n^{\epsilon'}}{\overline{\nu}_n^\epsilon} + 2n^{-C/2}.$$

To control the ratio $\overline{\nu}_n^{\epsilon'}/\overline{\nu}_n^\epsilon$, we further divide into two subcases.

Subcase 1.1 In this subcase, we assume that

$$e^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}} \ll \mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_\epsilon)e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\}. \tag{90}$$

We then have

$$\begin{aligned}
&\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon'})e^{-(1-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} \\
&\leqslant \mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon})e^{-(1-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} + Ce^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}} &(91)\\
&= \mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon})e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}e^{-(\epsilon-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} + Ce^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}} \\
&= o(1)\cdot\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon})e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} + Ce^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}} &(92)\\
&\ll \overline{\nu}_n^\epsilon. &(93)
\end{aligned}$$

Here, (91) holds since $\exp\{-(1-\epsilon')mI(\alpha_0,z_0)\} \leqslant 1$ and $\mathbb{P}_{H_0}(z_0 \in \mathcal{B}_{\epsilon'}\backslash\mathcal{B}_\epsilon) \leqslant \mathbb{P}_{H_0}(z_0 \notin \mathcal{B}_\epsilon) \leqslant C\exp\{-(1-\epsilon)\rho^2/(2\tau^2)\}$. In (92), the equality holds since $\epsilon > \epsilon'$ and $nI(\alpha_0,z_0)$ is bounded from below uniformly when $z_0 \in \mathcal{B}_\epsilon$ by a sequence that diverges to infinity. Finally, (93) holds since both terms in (92) are $o(\overline{\nu}_n^\epsilon)$ as $n \to \infty$ under (90). Hence,

$$\mathbb{P}\left\{\ell(\sigma,\widehat{\sigma}) > \overline{\nu}_n^\epsilon\right\} \leqslant \frac{\overline{\nu}_n^{\epsilon'}}{\overline{\nu}_n^\epsilon} + 2n^{-C/2} = o(1). \tag{94}$$

Subcase 1.2 In this case, we consider the situation complementary to (90), namely

$$\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_\epsilon)e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} \lesssim e^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}}. \tag{95}$$

Equation (95) leads to

$$\begin{aligned}
\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} &\leqslant \mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_\epsilon)e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} + \mathbb{P}_{H_0}(z_0 \notin \mathcal{B}_\epsilon) \\
&\lesssim e^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}}. &(96)
\end{aligned}$$

For the first term in $\overline{\nu}_n^{\epsilon'}$, we have

$$\begin{aligned}
&\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon'})e^{-(1-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} \\
&= \mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon'})e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}e^{-(\epsilon-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} \\
&= o(1)\,\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon'})e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} &(97)\\
&= o(1)\,\mathbb{E}_{H_0}^{\alpha_0,z_0}\left\{e^{-(1-\epsilon)\frac{n}{2}I(\alpha_0,z_0)}\right\} \\
&\ll e^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}}. &(98)
\end{aligned}$$

Here (97) holds since $nI(\alpha_0, z_0)$ is bounded from below uniformly when $z_0 \in \mathcal{B}_{\epsilon'}$ by a sequence that diverges to infinity and $\epsilon > \epsilon'$. The bound (98) is due to (96).

Under (95), we then have

$$\overline{\nu}_n^{\epsilon'} = \mathbb{E}_{H_0}^{\alpha_0, z_0}\left\{1(z_0 \in \mathcal{B}_{\epsilon'})e^{-(1-\epsilon')\frac{n}{2}I(\alpha_0,z_0)}\right\} + e^{-(1-\epsilon')\frac{\rho^2}{2\tau^2}} \ll e^{-(1-\epsilon)\frac{\rho^2}{2\tau^2}} \lesssim \overline{\nu}_n^{\epsilon}.$$

Hence, the desired bound (94) continues to hold.

Case 2 When

$$\overline{\nu}_n^{\epsilon} < n^{-(1+C/2)} < n^{-1}, \tag{99}$$

then

$$\begin{aligned}
\mathbb{P}\left\{\ell(\sigma, \widehat{\sigma}) > \overline{\nu}_n^{\epsilon}\right\} &= \mathbb{P}\left\{\ell(\sigma, \widehat{\sigma}) > 0\right\} \\
&\leqslant \sum_{i=1}^{n} \mathbb{P}(\sigma_i \neq \widehat{\sigma}_i) \\
&\leqslant n\overline{\nu}_n^{\epsilon} + 2n^{-C} \\
&\leqslant n^{-C/2} + 2n^{-C} = o(1).
\end{aligned}$$

Here, the second inequality is a union bound. The third inequality is due to (89) and the last inequality holds due to (99). This completes the proof.

C.3 Proof of Theorem 9

The lower bound can be established by adapting some arguments spelled out in Section 3 of Gao and Ma (2020). We include them below for the manuscript to be self-contained.

For any $0 < \epsilon_2 < \epsilon_1 < 1/2$, we have

$$\underline{\nu}_n^{\epsilon_1} \leqslant \underline{\nu}_n^{\epsilon_2}, \quad \frac{\underline{\nu}_n^{\epsilon_1}}{\underline{\nu}_n^{\epsilon_2}} \to 0.$$

Therefore, for any fixed $\epsilon \in (0, 1/2)$, we may choose a fixed $\epsilon' > 0$ and a sequence $\delta' = \delta_n'$ such that

$$\frac{1}{n} \ll \delta' \ll 1, \quad \delta' \underline{\nu}_n^{\epsilon'} \gtrsim \underline{\nu}_n^{\epsilon}. \tag{100}$$

Then, we choose a $\sigma^* \in [2]^n$ such that $n_u(\sigma^*) \in [(1-\delta')n/2, (1+\delta')n/2]$ for $u = 1, 2$. Let $\mathcal{C}_u(\sigma^*) = \{i \in [n] : \sigma_i^* = u\}$. Then we choose some $\widetilde{\mathcal{C}}_1 \subset \mathcal{C}_1(\sigma^*)$ and $\widetilde{\mathcal{C}}_2 \subset \mathcal{C}_2(\sigma^*)$ such that $|\widetilde{\mathcal{C}}_1| = |\widetilde{\mathcal{C}}_2| = \lceil (1-\delta')n/2 \rceil$. Define

$$T = \widetilde{\mathcal{C}}_1 \cup \widetilde{\mathcal{C}}_2, \quad \mathcal{Z}_T = \left\{\sigma \in [2]^n : \sigma_i = \sigma_i^* \text{ for all } i \in T\right\}.$$

The set $\mathcal{Z}_T$ corresponds to a sub-problem that we only need to estimate the clustering labels $\{\sigma_i\}_{i \in T^c}$.

Given any $\sigma \in \mathcal{Z}_T$, the values of $\{\sigma_i\}_{i \in T}$ are known. Now, we define the subspace

$$\mathcal{P}_n^0 = \left\{\mathcal{M}_n(\sigma, H, \mu, \tau, F_\alpha) \in \mathcal{P}_n : \sigma \in \mathcal{Z}_T\right\}.$$

We have $\mathcal{P}_n^0 \subset \mathcal{P}_n$ by the construction of $\mathcal{Z}_T$. This gives the lower bound

$$\inf_{\widehat{\sigma}} \sup_{\mathcal{P}_n} \mathbb{E}\ell(\sigma, \widehat{\sigma}) \geqslant \inf_{\widehat{\sigma}} \sup_{\mathcal{P}_n^0} \mathbb{E}\ell(\sigma, \widehat{\sigma}) = \inf_{\widehat{\sigma}} \sup_{\sigma \in \mathcal{Z}_T} \frac{1}{n} \sum_{i=1}^n \mathbb{P}(\widehat{\sigma}_i \neq \sigma_i). \tag{101}$$

The last equality above holds because for any $\sigma^1, \sigma^2 \in \mathcal{Z}_T$, we have $\frac{1}{n}\sum_{i=1}^n 1(\sigma_i^1 \neq \sigma_i^2) = O(\delta') = o(1)$ so that $\ell(\sigma^1, \sigma^2) = (1/n)\sum_{i=1}^n 1(\sigma_i^1 \neq \sigma_i^2)$. Continuing from (101), we have

$$\begin{aligned} \inf_{\widehat{\sigma}} \sup_{\sigma \in \mathcal{Z}_T} \frac{1}{n} \sum_{i=1}^n \mathbb{P}(\widehat{\sigma}_i \neq \sigma_i) &\geqslant \frac{|T^c|}{n} \inf_{\widehat{\sigma}} \sup_{\sigma \in \mathcal{Z}_T} \frac{1}{|T^c|} \sum_{i \in T^c} \mathbb{P}(\widehat{\sigma}_i \neq \sigma_i) \\ &\geqslant \frac{|T^c|}{n} \frac{1}{|T^c|} \sum_{i \in T^c} \inf_{\widehat{\sigma}_i} \operatorname*{ave}_{\sigma \in \mathcal{Z}_T} \mathbb{P}(\widehat{\sigma}_i \neq \sigma_i). \end{aligned} \tag{102}$$

For each $i \in T^c$,

$$\inf_{\widehat{\sigma}_i} \operatorname*{ave}_{\sigma \in \mathcal{Z}_T} \mathbb{P}(\widehat{\sigma}_i \neq \sigma_i) \geqslant \operatorname*{ave}_{\sigma_{-i}} \inf_{\widehat{\sigma}_i} \left\{ \frac{1}{2} \mathbb{P}_{(\sigma_{-i}, \sigma_i = 1)} (\widehat{\sigma}_i \neq 1) + \frac{1}{2} \mathbb{P}_{(\sigma_{-i}, \sigma_i = 2)} (\widehat{\sigma}_i \neq 2) \right\}. \tag{103}$$

Now consider any fixed pair $(\mathbb{P}_{(\sigma_{-i}, \sigma_i=1)}, \mathbb{P}_{(\sigma_{-i}, \sigma_i=2)})$. Let m_1 and m_2 be the number of nodes with label 1 and 2 in σ_{-i}, respectively. Let $\bar{m} = m_1 \vee m_2$. By the construction of $\mathcal{Z}_T$, we have

$$\left| \bar{m} - \frac{n}{2} \right| \leqslant \frac{\delta' n}{2}.$$

By data processing inequality, the total variation distance between this pair of distributions satisfies

$$\mathrm{TV}(\mathbb{P}_{(\sigma_{-i}, \sigma_i=1)}, \mathbb{P}_{(\sigma_{-i}, \sigma_i=2)}) \geqslant \mathrm{TV}(\mathbb{P}_{\bar{m}}^0, \mathbb{P}_{\bar{m}}^1), \tag{104}$$

where $\mathbb{P}_{\bar{m}}^0$ and $\mathbb{P}_{\bar{m}}^1$ refer to the null and the alternative distributions in (13) with $\bar{m}$ observations from either community. Continuing (104), we further obtain from Lemmas 5 and 6 that

$$\mathrm{TV}(\mathbb{P}_{(\sigma_{-i}, \sigma_i=1)}, \mathbb{P}_{(\sigma_{-i}, \sigma_i=2)}) \geqslant \mathrm{TV}(\mathbb{P}_{\bar{m}}^0, \mathbb{P}_{\bar{m}}^1) \geqslant \underline{\nu}_n^{\epsilon''}, \quad \text{for any } \epsilon'' \in (0, 1/2),$$

where we have used the second last display and the fact that $\delta' = o(1)$. Together with (101) and (102), this implies that for any $\epsilon'' \in (0, 1/2)$,

$$\inf_{\widehat{\sigma}} \sup_{\mathcal{P}_n} \mathbb{E}\ell(\sigma, \widehat{\sigma}) \gtrsim \delta' \, \underline{\nu}_n^{\epsilon''}.$$

We complete the proof by observing (100).

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pages 36–43. ACM, 2005.

Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, and Yichen Qin. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(1):8393–8484, 2017.

Moses Charikar, Sudipto Guha, Éva Tardos, and David B Shmoys. A constant-factor approximation algorithm for the k-median problem. *Journal of Computer and System Sciences*, 65(1):129–149, 2002.

Yudong Chen, Xiaodong Li, and Jiaming Xu. Convexified modularity maximization for degree-corrected stochastic block models. *The Annals of Statistics*, 46(4):1573–1602, 2018.

Rob Cross and Andrew Parker. *The hidden power of social networks: Understanding how work really gets done in organizations.* Harvard Business Review Press, 2004.

Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Chao Gao and Zongming Ma. Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science*, to appear, 2020.

Chao Gao, Zongming Ma, Anderson Y Zhang, and Harrison H Zhou. Achieving optimal misclassification proportion in stochastic block models. *Journal of Machine Learning Research*, 18(60):1–45, 2017.

Chao Gao, Zongming Ma, Anderson Y Zhang, and Harrison H Zhou. Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185, 2018.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, and Edoardo M Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.

Peter D Hoff. Random effects models for network data. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers.* Citeseer, 2003.

Peter D Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in neural information processing systems*, pages 657–664, 2008.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Jiashun Jin. Fast community detection by score. *The Annals of Statistics*, 43(1):57–89, 2015.

Varun Jog and Po-Ling Loh. Information-theoretic bounds for exact recovery in weighted stochastic block models using the renyi divergence. *arXiv preprint arXiv:1509.06418*, 2015.

Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

Pavel N Krivitsky and Mark S Handcock. Fitting latent cluster models for networks with latentnet. *Journal of Statistical Software*, 24(i05), 2008.

Pavel N Krivitsky, Mark S Handcock, Adrian E Raftery, and Peter D Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3):204–213, 2009.

Rafał Latała, Ramon van Handel, and Pierre Youssef. The dimension-free structure of nonhomogeneous random matrices. *Inventiones mathematicae*, 214(3):1031–1080, 2018.

Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

Zhuang Ma, Zongming Ma, and Hongsong Yuan. Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, 21(4):1–67, 2020.

Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, 10(9):e0136497, 2015.

Cristopher Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

Cosma Rohilla Shalizi and Dena Asta. Consistency of maximum likelihood for continuous-space network models I. *arXiv preprint arXiv:1711.02123*, 2021.

Amanda L Traud, Eric D Kelsic, Peter J Mucha, and Mason A Porter. Comparing community structure to characteristics in online collegiate social networks. *SIAM review*, 53(3): 526–543, 2011.

Amanda L Traud, Peter J Mucha, and Mason A Porter. Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012.

Haolei Weng and Yang Feng. Community detection with nodal information: Likelihood and its variational approximation. *Stat*, 11(1):1–17, 2022.

Yun-Jhong Wu, Elizaveta Levina, and Ji Zhu. Generalized linear models with low rank effects for network data. *arXiv preprint arXiv:1705.06772*, 2017.

Anderson Y Zhang and Harrison H Zhou. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.

Jingfei Zhang, Will Wei Sun, and Lexin Li. Network response regression for modeling population of networks with covariates. *arXiv preprint arXiv:1810.03192*, 2018.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. *SIAM Journal on Mathematics of Data Science*, 2(2):265–283.